Moshfeghi, Yashar (2012) *Role of emotion in information retrieval.* PhD thesis.

ROLE OF EMOTION IN INFORMATION RETRIEVAL

by

Yashar Moshfeghi

Submitted in fulfilment of the requirements for the title of Doctor of Philosophy

School of Computing Science
College of Science and Engineering
University of Glasgow

Glasgow, January 15, 2012

Abstract

The main objective of Information Retrieval (IR) systems is to satisfy searchers' needs. A great deal of research has been conducted in the past to attempt to achieve a better insight into searchers' needs and the factors that can potentially influence the success of an Information Retrieval and Seeking (IR&S) process.

One of the factors which has been considered is searchers' emotion. It has been shown in previous research that emotion plays an important role in the success of an IR&S process which has the purpose of satisfying an information need. However, these previous studies do not give a sufficiently prominent position to emotion in IR, since they limit the role of emotion to a secondary factor, by assuming that a lack of knowledge (the need for information) is the primary factor (the motivation of the search).

In this thesis, we propose to treat emotion as the principal factor in the system of needs of a searcher, and therefore one that ought to be considered by the retrieval algorithms. We present a more realistic view of searchers' needs by considering not only theories from information retrieval and science, but also from psychology, philosophy, and sociology. We extensively report on the role of emotion in every aspect of human behaviour, both at an individual and social level. This serves not only to modify the current IR views of emotion, but more importantly to uncover social situations where emotion is the primary factor (i.e., source of motivation) in an IR&S process.

We also show that the emotion aspect of documents plays an important part in satisfying the searcher's need, in particular when emotion is indeed a primary factor. Given the above, we define three concepts, called emotion need, emotion object and emotion relevance, and present a conceptual map that utilises these concepts in IR tasks and scenarios.

In order to investigate the practical concepts such as emotion object and emotion relevance in a real-life application, we first study the possibility of extracting emotion from text, since this is the first pragmatic challenge to be solved before any IR task can be tackled. For this purpose, we developed a text-based emotion extraction system and demonstrate that it outperforms other available emotion extraction approaches.

Using the developed emotion extraction system, the usefulness of the practical concepts mentioned

above is studied in two scenarios: movie recommendation and news diversification.

In the movie recommendation scenario, two collaborative filtering (CF) models were proposed. CF systems aim to recommend items to a user, based on the information gathered from other users who have similar interests. CF techniques do not handle *data sparsity* well, especially in the case of the *cold start* problem, where there is no past rating for an item. In order to predict the rating of an item for a given user, the first and second models rely on an extension of state-of-the-art memory-based and model-based CF systems. The features used by the models are two emotion spaces extracted from the movie plot summary and the reviews made by users, and three semantic spaces, namely, actor, director, and genre. Experiments with two MovieLens datasets show that the inclusion of emotion information significantly improves the accuracy of prediction when compared with the state-of-the-art CF techniques, and also tackles data sparsity issues.

In the news retrieval scenario, a novel way of diversifying results, i.e., diversifying based on the emotion aspect of documents, is proposed. For this purpose, two approaches are introduced to consider emotion features for diversification, and they are empirically tested on the TREC 678 Interactive Track collection. The results show that emotion features are capable of enhancing retrieval effectiveness.

Overall, this thesis shows that *emotion plays a key role in IR and that its importance needs to be considered*. At a more detailed level, it illustrates the crucial part that emotion can play in

- searchers, both as a primary (emotion need) and secondary factor (influential role) in an IR&S process;
- enhancing the representation of a document using emotion features (emotion object); and finally,
- improving the effectiveness of IR systems at satisfying searchers' needs (emotion relevance).

Acknowledgements

This thesis would not have been possible without the immense support that I received during the course of my PhD. I could not hope to do justice to this section given all the people who have contributed to this work. Nevertheless, I would still like to thank a number of people whose support made this long journey endurable.

My parents, with their unconditional love, support and encouragement made it possible for me to complete this work.

I would like to thank my supervisors, Joemon M. Jose and Keith van Rijsbergen who had enough faith in my skills to grant me this opportunity.

A great deal of gratitude must go to Benjamin Piwowarski, who was an inspiration; your feedback, supervision and encouragement have been invaluable.

I would especially like to thank Sachi Arafat and Alvaro Huertas, who commented on various drafts of this thesis.

The people with whom I have shared an office with over the last four years, such as Ali Azimi, Ingo Frommholz, Erik Graf have provided me with a stimulating research environment and camaraderie.

The assistance of David Hannah and Guido Zuccon with some of the experiments in this thesis are also appreciated.

Finally, I would like to thank the members of the Information Retrieval Group for their kind support and interest in my research.

Contents

List of Figures

List of Tables

Part I

INTRODUCTION

Chapter 1

Introduction

1.1 Introduction

This thesis investigates the benefits of considering emotion for the purpose of Information Retrieval (IR). While the role of emotion in human activities and behaviour has been well-studied in sociology, psychology, and philosophy communities, its role in Information Retrieval and Science (IR/IS) has not been investigated thoroughly. This thesis attempts such an investigation, and argues that (a) by considering emotion, IR can be founded on a more realistic understanding of the searcher and the search processes; (b) IR systems built upon this understanding can better address searcher's needs, and (c) this in turn can lead to more effective search systems.

Emotion is an important factor influencing every aspect of human behaviour, including rationality and decision-making [Iza91]. On the other hand, IR applications are pervasive, and used by human beings to fulfil daily social needs: entertainment, dating, getting to know people, maintaining friendship, gaming, etc. Therefore, it is natural to claim that emotion is a significant factor influencing the way IR applications are used. Prior research in IR/IS that considers emotion has undergone a number of significant changes, from focusing on the influence of emotion in information seeking behaviour [Kuh93] to postulating the potential of emotion as a relevance feedback mechanism [AAJ10]. Emotion-based works have been expanded to different domains of IR/IS, e.g., presenting efficient algorithms for extracting emotion from documents [SPI09], visualisation of the emotion aspects of documents [LSL03], and employing user mood to improve the recommendation of movies [WT10]. However, a global picture (structure) which relates prior works together from an IR perspective is missing. In this thesis, we attempt to reframe the role of emotion in IR/IS in a way that not only explains the relations between prior works in this area but also further explores the potential of emotion in information retrieval and seeking (IR&S). In this regard, this thesis investigates the nature of emotion and how it can be used to expand the fundamental concepts of IR, such as searcher needs, document representation, and relevance.

The rest of this chapter introduces the motivations, aims and achievements of the thesis, and finally presents the outline of the remainder of this thesis.

1.2 Motivation

The idea that *IR systems help searchers to overcome their information need (IN)* is a leitmotif since the early days of IR: the main task is to locate documents containing information *relevant* to such needs. Within this view, a searcher is understood as an agent that interacts with an IR system with the intention of seeking information [Kuh93]. The *information* can be defined as facts, propositions, and concepts, as well as evaluative judgements such as opinion [Wil93].

This standard (and dominant) view does not sufficiently consider the searchers' needs. Information science researchers have argued about the existence of needs other than IN, and discussed their roles in the cognitive aspects of human beings and in IR&S behaviour. Examples include Wilson's interrelation between physiological, affective and information needs in IR&S behaviour [Wil93], Kuhlthau's uncertainty principle [Kuh93], and Nahl's affective load theory [Nah05]; these studies have investigated the role of affective and cognitive experience of a searcher in an information seeking process model.

Although these views better capture the searchers' mind compared to the traditional view, their accounting for the role of emotion is limited to its relation with cognition in the process of satisfying an IN in an IR&S behaviour, e.g., Kuhlthau's [Kuh93] model. Therefore, emotion plays a marginal role in these views in their conception of human needs. For example, in an IR&S scenario, where searchers' task is to find documents that are topically relevant to a given query (e.g., Iraq War), the emotion that they experience during the completion of this task influences their performance and satisfaction. Another example is that of Arapakis et al. [AKJ09] that investigated the use of facial expressions and peripheral physiological signals as implicit indicators of topical relevance.

Others, e.g., Wilson [Wil93], consider a more autonomous role for affect and define *affective need* as an independent need which can motivate an IR&S behaviour. For example, gathering information to satisfy affective needs, such as the need for security, for achievement, or for dominance [Wil93]. However, there is no operationalisation of this *affective need* so that it could be used in real IR systems.

In general, the current landscape of the role of emotion in IR&S behaviour is incomplete. It is clear that people use computers for individual as well as social purposes, such as entertainment, dating, getting to know people, finding 'friends', gaming, etc., which strongly indicates that searchers try to satisfy other aspects of their needs than just information needs. The current views of emotion in IR/IS do not sufficiently explain these types of activities accurately, even though it is clear that searchers use emotionally-rich documents from the Internet to satisfy these needs.

In this thesis, we argue that human emotion is a central motivation (either directly or indirectly) behind all IR&S behaviour. This is justified by arguments beyond IR/IS, coming from domains such as philosophy, psychology, sociology, and communication. These domains describe the role of emotion and its importance in many aspects of human life. Based on the study of these domains, we define the concept of *emotion need* (Chapter 4). Such a conceptualisation allows human needs, in the context of IR&S, to be better encompassed.

The pervasiveness of emotionally-rich content on the web, such as movies, music, images, news, blogs, customer review, Facebook comments and Twitter, highlights the demand for such contents, and, indirectly, their role in satisfying searchers' needs. Although this issue has been investigated at length in other domains, in particular psychology of human behaviour, it has been relatively ignored in IR scenarios. Research in IR/IS is limited to looking at the emotion side of the documents in a traditional sense, which is to satisfy the IN of searchers. The influence of the emotion aspect of documents in satisfying emotion need is not covered by current research. In this thesis, we argue that it is important to consider the emotion aspect of documents to better understand and satisfy searchers' needs. This is done through developing a theoretical map centred on the notion of *emotion relevance*.

1.3 Thesis Statement

The statement of this thesis is that by considering emotion (a) IR models can be based on a more realistic understanding of the searcher and search processes; (b) that IR systems built upon this can better address searcher's needs and; (c) that this in turn can lead to more effective search systems. In order to do this, three fundamental concepts of IR are redefined: searchers' needs, document representation, and relevance.

we first discuss the role of emotions in human behaviour, focusing on the information seeking process (Chapter 2), and review other domains such as philosophy, psychology and sociology (Chapter 3). This allows a central concept of this thesis to be defined: that of emotion need. Its relation with other needs such as information need is investigated (Chapter 4).

We introduce the emotion object to emphasise the fact that documents can be represented by their emotion aspect analogically to their representation as information objects (Chapter 4). The feasibility of extracting emotion from textual documents is investigated to see how, in practice, emotions can be used by IR systems (Chapter 5). Having received positive results from this investigation (Chapter 5) we go on to define emotion relevance as a relationship between emotion objects and searchers' needs. Following Saracevic's terminology [Sar07], this is a new manifestation of relevance, called *emotion relevance*.

In order to illustrate how the emotion relevance can be used in practice, we chose to experiment with two IR tasks, namely collaborative filtering (CF) and diversity.

CF systems aim to recommend items to a user, based on the information gathered from other users who have similar interests. CF is interesting because people use computers for social as well as individual purposes. In particular, users are more and more dependent on the opinions and views of others in making decisions, and hence influence their IR&S behaviour; This is due to: (1) the social nature of humans and the culture shared by them, [Wil95] and (2) the pervasiveness of opinions and views online [GH06]. The IR research area that is tied with such a manifestation of the social nature of humans is *CF systems*. We show that the performance of the CF systems significantly improved when incorporating emotion features, in particular in the cases where data is sparse, i.e., only a few ratings are available (Chapter 6 and 7).

Diversity is interesting because searchers' queries are either ambiguous or multi-faceted [SJRS07]. A popular approach for search engines is to use the text content of documents for disambiguation [CG98]. In a preliminary study, we investigated the role of emotion aspect of documents in diversifying search results. The results provide empirical evidence that diversification of rankings using emotion representation of the documents can improve the effectiveness of the system (Chapter 8).

All the experiments conducted within this thesis demonstrate the practicality of just a small part of the emotion-based concepts defined above and, as a result, a great potential of research is introduced to the IR/IS community.

1.4 Research Objectives

Overall, this work is an exploration of a fairly new IR territory. The goal is to consider a more realistic understanding of the searcher and search processes in IR, based on emotion, to better address searcher needs and in turn to improve the effectiveness of search systems. In particular, the aim is to:

1. Study the current IR/IS views on the role of emotion in IR&S behaviours.
2. Study the role of emotion in the personal and social life of human beings in the philosophy, psychology and sociology literature.
3. Reframe the fundamental concepts in IR, such as searchers' need, document representation, and relevance, according to (1) and (2).
4. Introduce a conceptual map through which these concepts can be used in a structural/systematic way. This map can be used to describe a variety of IR scenarios.
5. Investigate the way the conceptual map can be employed to develop emotion-based IR systems, and the pragmatic challenges, the most critical of these being the effectiveness of extracting emotion from documents.
6. Investigate the use of the map in real-life applications involving emotion-rich content and behaviour; in this case, a study of the role of emotion in a movie recommender system and in a news diversification task (see 1).

1.5 Publications

A number of publications have been written during the course of this thesis:

1. Yashar Moshfeghi, Joemon M. Jose: Role of Emotional Features in Collaborative Recommendation. ECIR 2011, pages 738–742.
2. Yashar Moshfeghi, Guido Zuccon, Joemon M. Jose: Using Emotion to Diversify Document Rankings. ICTIR 2011, pages 337–341.
3. Yashar Moshfeghi, Benjamin Piwowarski, Joemon M. Jose: Handling data sparsity in collaborative filtering using emotion and semantic based features. SIGIR 2011, pages 625–634.
4. Ioannis Arapakis, Yashar Moshfeghi, Hideo Joho, Reede Ren, David Hannah, Joemon M. Jose: Enriching user profiling with affective features for the improvement of a multimodal recommender system. CIVR 2009, pages 29–37.
5. Yashar Moshfeghi, Deepak Agarwal, Benjamin Piwowarski, Joemon M. Jose: Movie Recommender: Semantically Enriched Unified Relevance Model for Rating Prediction in Collaborative Filtering. ECIR 2009, pages 54–65.
6. Ioannis Arapakis, Yashar Moshfeghi, Hideo Joho, Reede Ren, David Hannah, Joemon M. Jose: Integrating facial expressions into user profiling for the improvement of a multimodal recommender system. ICME 2009, pages 1440–1443.
7. Yashar Moshfeghi: Affective adaptive retrieval: study of emotion in adaptive retrieval. SIGIR 2009, page 852.

1.6 Overall Layout and Outline

This thesis is divided into four parts and ten chapters:

Part I: Introduction (Chapters 1 – 3) This part provides background and motivation. Chapter 2 further elaborates on motivation, and defines fundamental concepts in IR/IS by briefly discussing the state-of-the-art views of relevance and IR&S models, focusing on the role of emotion. Chapter 3 introduces the nature of emotion and its explanatory models, as well as its role in individual behaviour such as rationality, decision-making, attention, and memory retrieval. Its role in the social aspects of human life is then discussed through the concepts of communication, empathy, coping and finally entertainment theories.

Part II: Theoretical Contribution (Chapter 4) Chapter 4 introduces the main concepts defined in this thesis, namely emotion need, emotion object and emotion relevance. It defines a conceptual map in which these concepts are employed to better satisfy searchers' needs.

Part III: Practical Contribution (Chapters 5 – 8) In this part four practical contributions are presented. Chapter 5 presents a systematic analysis of current text-based emotion-extracting techniques. The best performing of these is used for the remaining experiments. Chapter 6 studies the role of emotions extracted from movie reviews and plot summaries to improve the accuracy of a memory-based CF system. Chapter 7 further investigates their role in a more elaborate CF system in order to overcome two important issues in such systems, data sparsity and cold start problems. Finally, Chapter 8 studies the effectiveness of using the emotion representation of documents (news articles) to diversify ranking results in order to better cover relevant subtopics.

Part IV: Conclusion (Chapter 9) Chapter 9 describes the contributions of this thesis, the conclusions drawn from the experiments in Part III, and the thesis overall, and describes avenues for future work.

Chapter 2

Background and Motivation

2.1 Introduction

This chapter provides the background for the concepts used in this thesis, and describes both the research and application context within which this work is situated. It begins with an introduction to IR in Section 2.2, and deals with the central concept of relevance in Section 2.3. This is important to define emotion need, which in turn is necessary to define a more realistic user model. In Section 2.4, information seeking behaviour (as a dynamic point of view) is studied to understand the role of emotion in the current views on the information seeking process.

2.2 Information Retrieval

The term *Information Retrieval (IR)* was coined by the mathematician and physicist Calvin N. Mooers (1919–1994). He defined IR as the domain that *"embraces the intellectual aspects of the description of information and its specification for search, and also whatever systems, technique, or machines that are employed to carry out the operation"* [Moo50]. A similar definition can be found in [BYRN99, SM86], where the field of IR is defined as comprising representation, storage and access to information.

There are several ways to search and present information for a searcher, which are based on different search paradigms: (1) Ad-hoc IR, where a system returns an ordered list of documents in response to user information need represented as a set key terms. (2) Classification, where a document is associated with one or more classes. This can be used, for example, to build a hierarchy (either manually or automatically) so that users can browse into more and more specific categories until they find relevant information. (3) Clustering, where documents are grouped into coherent sets. The *coherence* depends on the task at hand, and can be used to group Web search results into sets that each answer one aspect of the query. (4) Filtering, where documents are selected from a stream of documents to be presented to the user. Such systems can be used to reduce the number of documents a user consults from, for example, an RSS source. (5) Recommendation, where documents (e.g. movies, songs, products) are presented to the user based on its profile and/or its past ratings.

Ad-hoc IR is one of the best-known ways of searching for information on the Web. It is the prototype of all other IR tasks because it includes all the concepts and ideas present in other tasks in some way. Hence, we describe it in order to give an overview of what the important components in IR are in general. At an abstract level, an ad-hoc IR process can be divided into following steps: (1) searchers submit their queries as a representation of their information need (IN); (2) the IR system matches searchers' queries with the representation of the documents in order to calculate document relevance scores; (3) documents get ranked based on their relevance score, calculated by the matching function, and those which have the highest scores are presented to the searchers; (4) searchers evaluate the retrieved documents and if they do not satisfy searchers' IN, searchers reformulate and resubmit their queries. This process can be repeated until searchers' IN is satisfied.

Documents can be relevant to searchers for different reasons. The most standard one is the similarity of the content of the document (i.e., topic) to the submitted query. This type of relevance is known as topical relevance, and is the relationship between the topics presented by the documents. Therefore, the way to capture this topical relevance depends on an IR model.

A variety of models have been proposed in IR, and are based on vector space, logic, and/or probability. Each has different strategies on *"making an assertion about the relevance of a document to a query"* [Jos98]. All these models require three components to be detailed: (1) a representation of documents, (2) a representation of searchers' IN, and (3) a matching function to assess the relevance based on (1) and (2).

In order to create document representation, an IR system requires a pre-processing step called indexing to be carried out, in which it extracts the features of document to characterise it. The indexing process is introduced due to the nature of documents, which are either unstructured, referring to documents which do not have clear, semantically overt, easy-for-a-computer structure[1] (e.g., news, journals, etc.), or semi-structured, referring to documents with explicit markup (e.g., XML files or the coding underlying webpages), in contrast to relational database which deals with structured data [MRS08].

After the indexing process, each document is represented by a set of keywords, each associated to a value that corresponds to the importance of the keyword in characterising the document. The indexing process can involve an additional process called normalisation where, for example, commonly occurring terms in textual documents are removed (known as stop-words removal) and some terms are conflated (i.e., stemming). A similar procedure applies to the submitted queries.

Once the representation of documents and/or queries has been chosen, IR systems have to compute how closely a document matches a user's query or request. This is usually reduced to the computation of a number (a score or a probability of relevance), which is used in turn to order documents to be presented to a user. Matching functions depend on the notion of relevance they employ.

2.3 Relevance

Relevance is a fundamental concept in IR, and a huge body of research exists that attempts to understand this concept so as to operationalise it for IR systems. A prominent view [Sar07] from information science defines it as *"a relation between information or information objects (the Ps) on the one hand and contexts, which include cognitive and affective states and situations (information need, intent, topic, problem, task; the Qs) on the other hand, based on some property reflecting a desired manifestation of relevance (topicality, utility, cognitive match; the Rs)."* [Sar07]

There are two main research problems associated with relevance: (1) notions relating to human judgement, such as relevance, are hard to grasp and these are hard to define for the purpose of

[1] Although some researcher suggest that there is no such a thing as unstructured data due to latent linguistic structure of human languages [MRS08] (p. 1).

automatic interpretation by a system; and (2) although there is some agreement about the concept of *relevance* per se, there is disagreement on *what should be considered as relevant* – probably since it depends on individual preferences [Sar07]. As a result, different definitions and models of relevance have been presented in the past, each tackling this problem from a particular point of view. The vast variety of relevance-motivated researchers (including Saracevic [Sar07] and Mizzaro [Miz97]) presents a framework within which the different variations of relevance are captured and integrated into a bigger picture.

The two most important approaches to relevance are those of Saracevic [Sar07] and Mizzaro [Miz97]. We now describe both approaches, focusing on those concerned by affect, emotion and mood. Saracevic considers different definitions and models of relevance as manifestations of relevance, whereas Mizzaro defines them as a relationship between four dimensions: system, user, context, and time.

2.3.1 Saracevic's View on Relevance

Saracevic proposed one of the initial frameworks for relevance in 1975 [Sar75]. Relevance is considered as a phenomenon, and he states that it should not be studied as the answer to a *"what is it?"* question, as a formally defined entity, but as an empirical one, similar to those in natural sciences. His work then tries to answer the fundamental question *"what is the nature of relevance?"* by investigating further detailed questions such as *"what are the manifestations, behaviour and effects of relevance?"* [Sar07]. The explanation for having different manifestations of relevance is:

> "because information science deals with creation and derivation, systems and users, we understood early on that there is not only one kind of relevance, but several. They were even labeled differently, like topical relevance, user relevance, and so on..." (p. 1919) [Sar07]

In total, five manifestations of relevance are described: system (or algorithmic), topical (or subjective), cognitive, situational, and affective relevance [Sar07]. Among these, his notion of affective relevance is particularly important for this study, as it is related to emotion. Affective relevance is defined as a *"relation between the intents, goals, emotions, and motivations of a user, and information (retrieved or in the systems file, or even in existence)."* [Sar07] Saracevic argues that *"affective relevance underlies other relevance manifestations, particularly situational relevance"* [Sar07]. His definition, in fact, is inspired by Kuhlthau's information seeking process model [Kuh91] (see Section 2.4.1).

2.3.2 Mizzaro's View on Relevance

Another important relevance framework is proposed by Mizzaro. In contrast to Saracevic [Miz97], Mizzaro argues that relevance can be conceptualised as a relation between four entities:

1. system: surrogate, document, information;
2. user: query, request, information need, problem;
3. context: topic, task, context, and each combination of these;
4. time: the various instances in the time from a problem arising until its solution.

Mizzaro's view limits the possible relevances that can be defined given these dimensions. Mizzaro defines information as *"a small bit of knowledge"*, where that knowledge exists only inside the agents (i.e. human) [Miz96]. Defining information in IR/IS is as difficult as defining relevance [Wil93], and the general approach to it is similar to Mizzaro's approach:

> "I assume that knowledge exists only inside the agents' [Knowledge States] KSs (thus a book does not contain knowledge). A datum is an entity of the physical world that, once perceived by an agent, leads to a noninferential transition of a KS that changes, say, from K^I to K^F [I stand for Initial and F stands for Final]. When this happens, the datum is said to carry information." [Miz96]

Although this framework seems comprehensive, we believe there is a significant omission, that of emotion. This limitation is particularly apparent from Mizzaro's definition of *information need* as the need for information considered as the *difference* between two states of (searchers') knowledge

[Miz96]. In this view, the use of an IR system will be limited to the cognitive side, leaving no ground for the emotion side of the user. Saracevic's definition of information is also unclear, as to whether it encompasses the emotion aspect of the document, and there is no (clear) definition of emotion need.

Despite being an advance compared to the current works in IR, in both Saracevic's and Mizzaro's approaches, emotion is at most a secondary factor. Since emotion is merely a secondary motivational factor, the research on the role of affect in the information seeking behaviour process is still based on the understanding of emotion that limited the two views (explained in Section 2.3.1 and 2.3.2). However, as we will see later, it is clearly important not only to understand the role of emotion in IN satisfaction, but also to understand and embrace new dimensions of using IR systems, especially those which directly relate to satisfying other needs, such as emotion needs (see Section 4.2).

Emotion plays an important role in every aspect of human life, such as decision-making, attention, and rationality; in scenarios where the searcher's main motivation is to engage in IR&S, the motivation is emotional rather than informational. Next section describes information seeking models where emotion has a more important role.

2.4 Information Seeking Theories

The previous section investigated relevance, its characteristics, and the issues with its definition by focusing on its conceptualisation in the two major IR frameworks. The following discusses information seeking models that specifically investigate or propose the role of emotion aspect of a user in an information seeking process: in particular, it presents Wilson's argument about the relation between needs.

There has recently been an increase in the amount of research on the role of affect in the information seeking behaviour process; particularly those influencing a searcher's cognitive operation [Kuh93, Ing93]. This section reviews research on aspects of information seeking behaviour that may influence the relationship between affect and user interaction.

First, a general overview of information seeking behaviour models is given, focusing on models considering the affective aspect of searchers in IR&S behaviour, such as Kuhlthau's information seeking process model [Kuh93], Wilson's information seeking behaviour [Wil93], and Nahl's affective load theory [Nah05]. These models are discussed from the perspective of their understanding on the role of emotion in IR&S.

There are many theories and models that attempt to explain information seeking behaviour. In general, they assume that complex IR&S behaviour can be broken down into a relatively small number of activity stages [FEM05] (p. 138).

Ellis [Ell93] proposes a model that considers the following phases: starting, chaining, browsing, differentiating, monitoring, extracting, verifying and ending. Marchionini ([Mar95], p. 49–60) proposes eight parallel sub-processes: (1) recognising an information problem (2) defining and understanding the problem (3) choosing a search system (4) formulating a query (5) executing search (6) examining results (7) extracting information (8) reflecting/iterating/stopping.

2.4.1 Kuhlthau's Information Seeking Process Model

Kuhlthau's *information seeking process* model is one of the first and most popular models to investigate the affective along with cognitive and physical aspects of a searcher in an information seeking process. Similar to Ellis' and Marchionini's models, Kuhlthau's model illustrates information seeking activities throughout a search session rather than at a given point in time. She proposes that people's feelings, thoughts and actions interact within their information seeking process. Kuhlthau's information seeking process model describes the searchers' common patterns of seeking meaning from information, to extend their knowledge state on a complex problem or topic which has a discrete beginning and ending [Kuh93]. As shown in Figure 2.1, her model divides an information seeking process into six steps: initiation, selection, exploration, formulation, collection, and presentation. Each step encompasses the affective, cognitive, and physical human experience. For example, in the initiation step, searchers experience feelings of uncertainty and vague thoughts. In the formulation phase, the searchers tend to experience more focused and clearer thoughts.

The fundamental principle behind Kuhlthau's information seeking process is the *uncertainty prin-*

	Initiation	Selection	Exploration	Formulation	Collection	Presentation	Assessment
Feelings (Affective)	Uncertainty	Optimism	Confusion Frustration Doubt	Clarity	Sense of direction/ Confidence	Satisfaction or Disappointment	Sense of accomplish-ment
Thoughts (Cognitive)	vague			focused	increased	interest	Increased self-awareness
Actions (Physical)	seeking	relevant Exploring	information	seeking	pertinent Documenting	information	

Figure 2.1: The Kuhlthau Model of Information Search Process [Kuh04] (p. 82)

ciple [Kuh93]. This refers to the existence of a cognitive state which causes feelings of anxiety and lack of confidence. Feelings of doubt, anxiety and frustration are in association with vague and unclear thoughts. The model shows that during a typical information seeking process, the thoughts of a searcher become clear and consequently their confidence increases and their feeling of doubt, anxiety and frustration decrease (see Figure 2.1). Kuhlthau determined that in the exploration state there is another sharp increase in uncertainty showing that information does not necessarily reduce uncertainty. In certain situations, it may even increase it. There will be scenarios in which searchers can be unsuccessful in finding what they seek.

Kuhlthau describes the beginning of the information seeking process as follows:

> "In the first stage, initiation, a person becomes aware of a gap in knowledge or a lack of understanding, where feelings of uncertainty and apprehension are common. At this point, the task is merely to recognise a need for information." [Kuh93] (p. 343)

Although this model is an important step towards understanding the role of emotion in IR/IS, it does not encompass many important aspects of emotion in IR. Kuhlthau considers emotion/affect as a factor influencing the information seeking process, rather than a need in itself.[2]

[2] Since Saracevic's definition of affective relevance is based on Kuhlthau's model, the same limitations apply in his work.

Moreover, Kuhlthau's model is limited by making uncertainty central, i.e., as driving the seeking process while we know (see Section 2.4) that positive or negative emotion states, high or low arousal level, such as stress or boredom respectively, could also motivate users to engage in an information seeking behaviour. Therefore, a key limitation lies in the fact that the affective side of searchers is interpreted as only being a secondary motivational source for information need. In this thesis, we consider emotion in a way that also deems it as a separate need.

For Kuhlthau, an emotion need is a need to change negative feelings caused by uncertainty during the initiation phase (e.g. feelings of doubt, anxiety and frustration) to feelings of satisfaction and comfort. This means that accompanying any IN is a more fundamental requirement to satisfy that need; and this fundamental requirement to *feel satisfied* is an emotional need. This need to be satisfied through changing negative feelings into positive ones through particular behaviours is known in psychology as *coping* (Section 3.5.3). Since the reason for discomfort is a lack of knowledge, searchers, via engaging in an information seeking process, can overcome their information need and consequently their emotion need. According to coping theory, one of the strategies to reduce negative feelings of anxiety and stress is to obtain information that can be used to reduce or diminish the cause of these feelings. Nahl explicitly referred to coping theory to develop her theory of affect.

2.4.2 Nahl's Affective Load Theory

Time is an important factor in information seeking processes and Kuhlthau's model does not consider it, e.g., the effect of a deadline on information seeking. Nahl took the uncertainty principle proposed by Kuhlthau and developed a theory in which the relation between affect and cognition is considered based on both *uncertainty* and *time pressure*.

The *affective load theory* [Nah05] of Nahl is an influential theory which presents a social-behaviour perspective on cognition and affect of searchers during an information seeking process. Nahl identifies three social-behaviour principles: (1) that *behaviour* is a function of both the cognitive and affective faculties of a user; (2) that cognitive behaviour is initiated, maintained and terminated by affective behaviour; and (3) that affective and cognitive behaviour operation systems are respectively binvalent and multivalent (bipolar and multiscale). Given these principles, she proposed that all information seeking behaviour involves affective states. It is important to create a connection between each cognitive behaviour and its affective support [Nah05].

From an operational point of view, affective load is defined as *"uncertainty multiplied by felt time pressure"* [Nah04]. Nahl defined uncertainty as the combined degree of irritation, frustration, anxiety and rage. High affective load indicates an ineffective cognitive behaviour. Once such a situation is identified, coping assistance can be provided to alleviate the disruptive affective states to achieve task success [Nah05].

There are, however, emotional needs other than the ones previously mentioned which should be satisfied and are not necessarily generated because of uncertainty. Instead, they can be caused by an emotion need such as the need to maintain a happy state or alleviate a sad state (which is not caused by uncertainty) or overcoming a low (i.e. bored) or high (i.e. stressed) arousal level. In these scenarios, explained further in Section 3.5.4, another aspect of information objects, the emotional aspect, satisfies the need of the user. Neither Kuhlthau's nor Nahl's models include this emotional aspect. To do so requires emotion need to be treated differently from simple IN.

2.4.3 Wilson IS Behaviour Model - The Separation of Needs

Wilson is another pioneer in this area whose work on information seeking behaviour is one of the most cited in this domain. He proposed the idea that there is more than one type of need for searchers. Wilson considers three types: physiological, affective, and informational. He argues that they are interrelated [Wil93], and that the recognition of a need in general (shown in Figure 2.2) is what makes a user engage in an information seeking behaviour. Based on his model (shown in Figure 2.2), factors influencing needs are at the personal level: physiological, affective, and cognitive needs shown in the person box. Each of these needs can directly motivate a searcher to engage in information seeking behaviour.

> "It will be quickly recognised that these three [need] categories are interrelated: physiological needs may trigger affective and/or cognitive needs; affective needs may give rise to cognitive needs; and problems relating to the satisfaction [of] cognitive needs (such as a failure to satisfy needs, or fear of disclosing needs) may result in affec-

tive needs (for example, for reassurance). These interrelationships are shown in 2.2 [adopted the figure name] which suggests that, as part of the search for the satisfaction of these needs, an individual may engage in information-seeking behaviour. Indeed, it may be advisable to remove the term ‘information needs’ from our professional vocabulary and to speak instead of ‘information-seeking towards the satisfaction of needs’.” [Wil93]

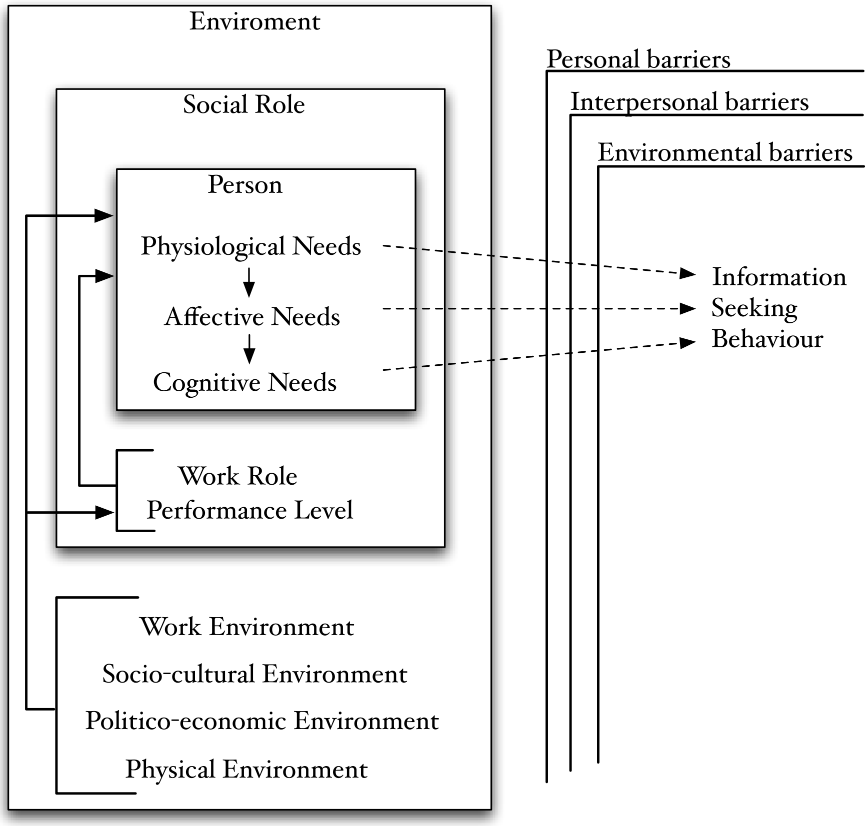


Figure 2.2: Factors influencing needs and information seeking behaviour [Wil93]

This description of the interplay between needs is particularly interesting and relevant to this thesis. Wilson also explains that the recognition and satisfaction of affective needs may not necessarily be a conscious experience. This argument is also made by many psychologists, such as Zillmann [Vor03] in his mood management theory (see Section 3.5.4 regarding needs in regular human activity). We adopt this view, as, although all information needs stem from an emotion need, a searcher may not be consciously aware of this underlying need (see Section 4.2).

Another interesting aspect of Wilson's work is the explanation of the existence of an important problem in information retrieval in general: the ambiguity in the definition of the *information* concept. He argued that this ambiguity lies not only in its numerous definitions, but also in the inappropriateness of the provided definitions. He discussed the fact that this problem gets even more complicated when there is (or not) a differentiation between facts, advice and opinion. Advice and opinion have value judgements, whereas fact does not [Wil93]. Wilson believes that the vagueness in definition of the concept of information, explained earlier, is one of the main problems causing difficulty in understanding and defining *information need* [Wil93].

The significant point in Wilsons' arguments is his emphasis on the importance of the separation that should be made between evaluative judgements (e.g. opinion) and non-evaluative judgements (e.g. facts), and its effect on the definition of a need. However, he argues that facts are also not free of evaluative judgements. This thesis differentiates between the facts (conceptualised as topicality) presented by an information object and the emotions which can be extracted from document (see Section 4.3).

2.5 Chapter Summary

This chapter provides the background for the concepts described in this thesis, and describes the context in which the work is situated. Current views in IR/IS on concepts such as relevance, information seeking behaviour, information need and associated problems are explained.

Although emotion need is mentioned in the literature, it has been ignored by the community, and by the majority of the work focused on the study of the satisfaction of searchers' information need. When mentioned, the role of emotion in IR&S behaviour is limited to a secondary factor that influences searchers' behaviour (either directly or through its effect on cognition) in satisfaction of IN.

The current definition of information in IR/IS does not clearly encompass the emotion side of documents and its effect in a searcher's mind. Facts and evaluative judgment are considered to be similar, and used to satisfy information need of searchers.

Regardless of considering emotion as a primary or secondary factor in IR&S behaviour, it is important to be able to extract emotion. This can be done in two ways: (1) from searchers during the IR&S behaviour through their physiological responses [SCKP08] and/or facial expressions [AMJ+09], or (2) from documents [AS05]. In this thesis, we are not interested in (1), and focus on the extraction of emotion from documents.

In order to broaden the view within the IR/IS domain of the role of emotion in IR&IS behaviour, we first study, in the next chapters, human behaviours in which emotion manifests itself the most, and then investigate whether any of these behaviours can be observed in the IR&S behaviour. For this purpose, the role of emotion in rationality, decision-making, and attention, as well as conceptualisation and communication of emotion, empathy and coping, are discussed in the next chapter.

In this chapter, we argued that

1. *Emotion need is prior to information need*, and
2. The purpose of the information seeking behaviour is to alleviate emotion need.
3. The facts (conceptualised as topicality) and the emotion which can be extracted from a document should be differentiated.

Chapter 3

Emotion in Social Sciences

3.1 Introduction

There is a large body of work regarding the nature and importance of the emotions, and a variety of contradictory theories have been proposed. Some maintain that the concept of emotion is unnecessary for the science of human behaviour [Duf62, Lin64]. Others, such as Izard [Iza71, IBM91] and Tomkins [Tom63], explain that emotion forms the primary motivational system of human beings. Some have said that emotions are only transient phenomena, while others believe that emotion is an intrinsic part of human lives [Sch59]. While some believe that emotion disrupts and disorganises behaviour, and is therefore a source of human problems [Arn60, Laz91, You61], others argue that emotions are organising and motivating, and serve to sustain behaviour [Iza71, Tom63, Sch59, Rap06].

This thesis follows the view that emotion systems motivate and organise behaviour, in particular IR&S behaviour. Therefore, it is important to know the difference between emotions, both in how it feels to experience them and in how people are affected by them [Iza91]. Section 3.2 gives an ontology of the theories of emotion. Terminologies associated with emotion and their differences are discussed in Section 3.3. In order to understand the role of emotion in IR&S behaviour, the role of emotion is considered at two levels; namely the individual (personal) and social levels.

At the individual level, the role of emotion in rational decision-making, attention and memory retrieval is discussed. At the social level, concepts such as communication of emotion, empathy, coping, and entertainment-based behaviour are discussed.

3.2 An Ontology of Emotion Theories

This section discusses the main groups of emotion theories, namely feeling and cognitive theories. For each group, we first define the general idea, then explain the most prominent theory of this group and discuss relevant criticisms. These different theories help to illustrate how the emotion underlying human behaviour can be framed in IR&S processes. Understanding the nature of emotion, how it is generated, and what it represents, is crucial for capturing and using it in IR&S behaviour models. We begin with feeling theories since they were the first group of theories proposed for emotion.

3.2.1 Feeling Theories

The conscious feeling of bodily changes is essential to consider having an emotional experience according to feeling theories. One of the most famous of these theories is the William James[1] feeling theory, also known as James-Lang theory[2]. According to this theory, emotion is a conscious feeling of bodily changes and nothing more than this. This view is explicitly expressed in his work:

> "My thesis ... is that the bodily changes follow directly the PERCEPTION of the exciting fact, and that our feeling of the same changes as they occur IS the emotion. (p. 189) ... If we fancy some strong emotion, and then try to abstract from our consciousness of it all the feelings of its characteristic bodily symptoms, we find we have nothing left behind, no 'mind-stuff' out of which the emotion can be constituted, and that a cold and neutral state of intellectual perception is all that remains. (p. 193)" [Jam84]

[1] William James was a prominent US philosopher of the late 19^{th} century

[2] C.G Lang was a Danish psychologist who published much the same idea at the same time

According to these theories, emotion is nothing more than physiological changes. Works that detect emotion from such changes can be categorised under these theories, e.g., Ekman's emotions extraction based on facial expression [ED94].

Feeling theories are criticised as it is debatable whether feelings are *necessary* or *sufficient* for/as emotional experiences. With regard to the argument about the necessity of feeling of bodily changes, critics argue that an exact set of bodily changes does not happen every time an emotion is experienced. Moreover, not all emotions have feelings of bodily changes. Examples of this include long-term emotions (e.g. love), or moral, intellectual and aesthetic emotions [Mil98]. As to sufficiency, Cannon[3] [Can27], Schachter and Singer [SS62] conducted a set of experiments and their results contradict the idea that feelings are sufficient for an emotional experience.

The other line of criticism is concerned with the concept of intentionality. Errol Bedford argued that the James-Lang theory fails to account for this particular feature of emotion: *"Emotions are about particular objects, and thoughts about such objects would seem to be a necessary feature of the emotions. For instance, pride involves some connection to me [self]; indignation implies unfairness; hope implies a desirable prospect. This brings the idea of cognition – in the form of belief – into the picture. This is because beliefs about material and formal objects seem to be essential to emotional experience. The main argument of Bedford was that emotions are to be distinguished, not in virtue of how they feel, but in virtue of their intentional object, which provide logical constraints on whether or not something is an emotion."* [Bed56]

Emotion-based experiences thus appear to have *more content* than any *associated feeling of bodily changes*, often not being associated at all with feelings, and appear to be particularly directed, i.e. the emotion tends to be caused by something specific. Emotions seem to be characterised by theories that further consider cognitive aspects.

3.2.2 Cognitive Theories

The content of mental (or cognitive) states is described as *thoughts* or *beliefs* which are often associated with values. Cognitive states change through processes of judgement (or *evaluative*

[3] Walter Cannon was a physiologist who attacked the James-Lang theory in his book named Bodily changes in Pain, Hunger, Fear and Rage in 1927.

thinking), e.g., *judging the truth of a belief*, by which the values associated with belief (and hence the beliefs themselves) are modified. Although some scientists equate emotions with cognitive mental states (known as pure cognitive theories approach), others require the presence of extra-cognitive features such as feelings. The latter is known as hybrid cognitive theories approach.

Judgementalism Cognitive theories of emotion revolve around describing: (a) the way by which an emotion appears in a mental state, and; (b) the nature of its association with the contents of the state (i.e. belief). In order to further study this group of theories, we focus on one of the best-known, which is of type (a), and is known as *Judgementalism*. For Judgementalism, an emotion is identical to an evaluative judgement. Robert Solomon [Sol73], one of the pioneers of Judgementalism, explained:

> "A change in what I am angry 'about' demands a change in my anger . . . if emotions were feelings, it would be a peculiar coincidence that the feelings were so faithful to our views of our situation. (p. 24) . . . I am angry at John' for taking . . . my car entails that I believe that John has somehow wronged me. The (moral) judgement entailed by my anger is not a judgement about my anger . . . My anger is that judgement. . . . to have an emotion is to hold a normative judgement about one's situation. (p. 27)" [Sol73]

The cognitive theory of emotion is particularly suited to the subject of this thesis. Searchers in an IR&S process are required to interact with a document (e.g., read the document), and judge its relevance. Since an IR&S process uses the concept of judgement (cognition), the emotion generated during this process seems to be a consequence of such mental activities. For example, Arapakis et al. [AJG08] show that there is a relation between emotion and relevance assessments of document.

Cognitive theories have been subject to much criticism, especially with respect to the plausibility of the argument which states that evaluative judgement is *necessary* for the emotion to occur. Critics argue that not all evaluative judgements are emotional, suggesting that something more than judgement is needed for emotion. An example illustrates this: *"many smokers believe smoking to be dangerous but smoke without fear. There are whole classes of evaluative judgements that are* *never likely to be the contents of emotions. (p. 29) . . . the judgement that Ashkenazy is a fine pianist . . . might give rise to envy or admiration or several other emotions, or perhaps to no emotion at all . . . Such cases suggest either that the evaluative beliefs associated with an emotion do not exhaust its content or that emotions are not individuated by their contents. (p. 29)"* [Gri89]

There is the possibility of conflict between emotions and the relevant evaluative judgements, e.g. fear of harmless objects, survivor guilt, irrational shame, etc., known as *recalcitrant* emotions. Moreover, emotions can precede conscious judgement; some emotions occur very quickly, while a process of conscious endorsement of an emotional appearance might take time. Finally, there is a peculiar emotional response to fiction and imagination which can give rise to anger, fear, sorrow, joy, etc. However, it is doubtful that we think that we are really in danger when afraid of a horror film.[4]

3.2.3 The proposed theoretical position for emotion-based IR

For the purpose of this thesis it is not crucial to know if an evaluative judgement, or indeed any cognitive activity, precedes (or accompanies) an affective reaction. Instead, it is clear from the above, both from the cognitive and feeling theories perspective, that *evaluative judgements are possible, that they possibly precede and/or accompany affective reactions*; and, further, that they often do.

In addition, for the situations we are concerned with, where users spend time looking at materials through interactions with a computer, phenomena such as decision-making (e.g. deciding about the relevance of documents), reading, comprehension and search, are all *cognitive phenomena*. This implies that any affective reaction is not merely *possibly* accompanied or preceded by *evaluative judgements*, but are *probably* so.

Given the above, cognitive theories are one of the foundations of this thesis. However, there can be affective reactions of a non-cognitive and non-perceptual type. In fact, it is plausible that if the user

[4]In response to such criticisms, perceptual theories as a branch of cognitive theories have been proposed which capture the advantages of judgementalism – in terms of accommodating the intentionality of emotion – without implying that the subject of 'recalcitrant' emotional experience is problematic. It thereby avoids the main issue in Judgementalism.

reacts in fear to certain signals, e.g. pictures of wars, over and over again, then it can perhaps be inferred, given this evidence, that: (a) they are socially conditioned this way (ontogenetic primacy); or (b) given further data about phylogenetic characteristics of people, that certain reactions (i.e. fear with respect to war , happiness with respect to pictures of children) are *norms*. However, the aspects of feeling theories which are included in an IR&S processes can be considered as special cases, and the argument about whether evaluative judgements are necessary or not in every case of affective reaction are not so relevant for the purpose of this thesis.

3.3 A Definition of Emotion

Given the theoretical position of this thesis (i.e., cognitive theories), we define terminologies associated with emotion. In this section, we start by defining the terminology associated with emotion in Section 3.3.1 and describe the proposed approaches for modelling it in Section 3.3.2.

3.3.1 Affect, Sentiment, Emotion, and Mood

The terms *affect, sentiment, emotion* and *mood* are often used interchangeably. However, there is a growing agreement between psychologists on their definition.

Ortony [Ort08] provides a clear definition of these concepts:

> "the term *affect* is used to refer to any aspect of mental life that has to do with value, examples include moods and preferences, emotions. *Affect* is a more general term than *emotion*. Affect always has one of two possible values, *and it is helpful to think of it as lying along a continuum from undifferentiated to differentiated, with vague positive and negative feelings [i.e. sentiment] being the most obvious examples of undifferentiated affect*. Specific discrete emotions are among the best examples of differentiated affect. One way to characterise the difference between affect and emotion is to view emotions as cognitively elaborated undifferentiated affect [ONR05] . . . . In this view, emotions (as opposed to undifferentiated affect) always involve cognition [CO00], even though these *cognitions* are not necessarily conscious." (p. 11–12)

There are other widely-accepted definitions other than Ortony's (above) which vary little from the above definition. Clore, Schwarz, and Conway [CSC94] define affect as being a positive or negative valence of an emotional experience, where affect and sentiment are considered to be the same concept. Frijda defines affect as *"the irreducible aspect [of emotion] that gives feelings their emotional, non-cognitive character"* [Fri93].

Following Ortony, we consider affect as a general (abstract) concept under which other concepts (e.g. sentiment, emotion, mood) are encompassed. We also take Clore et al.'s [CSC94] idea of the bivalency of sentiment by characterising it as a positive or negative feeling considered to be the central and most basic feature of emotion. We define emotion as cognitively elaborated sentiment, and mood as a general *pleasant* or *unpleasant* feeling (following Frijda's definition [Fri93]).

The difference between emotion and mood is that emotion has an immediate object or identifiable action or event whereas mood does not; and that emotion tends to be short and occur in bursts, whereas mood tends to persist longer and be flat [RSLW06].

3.3.2 Modelling Emotions

We discussed the relationship between the similar notions of affect, mood and emotions, and the relevance of the relation between affect and memory, for emotion-based approaches to IR. We now further develop the concept of emotion with a typology of emotions and models of their inter-relationships, i.e., what kinds of emotions there are and how they relate. There are two approaches to characterising emotion, namely dimensional and discrete[5] [CdRN09].

Dimensional Models

In the dimensional approach, emotion is considered as a continuous function of several dimensions. The best-known of such approaches is the *circumplex model* [Rus80], which defines emotions with respect to their coordinates in two dimensions, namely valence and activity. The valence dimension varies from positive to negative, and the activity dimension varies from passive to active. As shown

[5]There are numerous other categorisations but this is a common one. [GAT98]

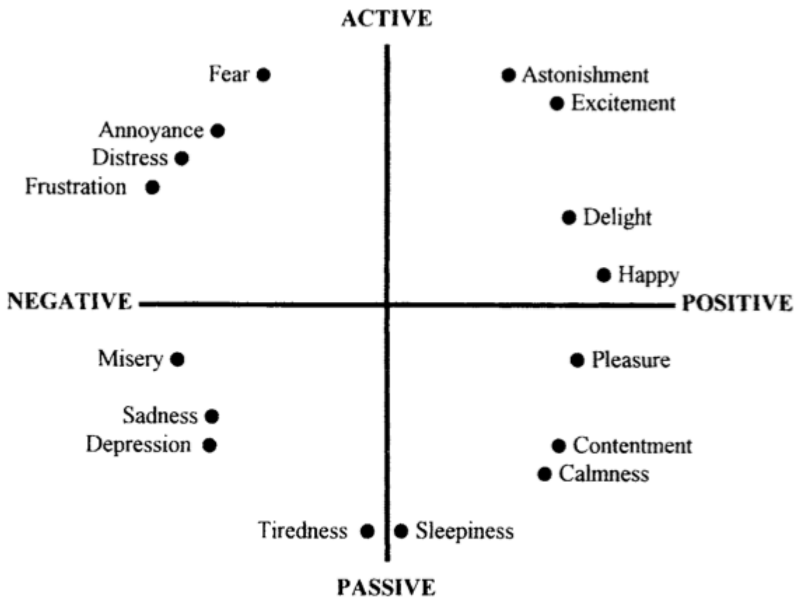


Figure 3.1: The circumplex model of emotions. (Adopted by Guerrero et al. [GAT98] [(p. 14)] from [Rus89]).

in Figure 3.1, these two dimensions are orthogonal and partition emotions into four quadrants.

Although this model is powerful in organising emotions [LD92], multiple criticisms are associated with it: The simplicity of the model in capturing the complex nature of emotion (e.g. anger and fear are very correlated in the model while they are two very distinct emotions) [LD92], and presence of non-emotional situations such as sleepiness and tiredness [SSKO87].

Discrete Models

Discrete approaches work on the idea that *"individuals experience basic emotions as distinct from one another"* [GAT98], i.e. there is no underlying conceptual space in which they can be placed. In this view, emotions are either basic or non-basic. Basic emotions are associated with distinct features, which may be biological, physiological, or semantic primitives. Non-basic emotions are defined as a *blend* of the basic emotions. For example, a combination of anger and fear could lead to the non-basic emotion of rage. A combination of sadness, surprise, and disgust could result in the non-basic emotion of disappointment [GAT98]. Basic emotions are universally understood, whereas the non-basic emotions are culturally dependent [EO79, Iza71, GAT98].

Table 3.1: Discrete Approaches in defining basic emotions

Theorist	Feature	Basic Emotions
Arnold [Arn60]	Cognitive	anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, sadness
Ekman [Ekm99]	Physiological	happiness, sadness, fear, anger, disgust and surprise
Frijda [Fri87]	Cognitive	desire, happiness, interest, surprise, wonder, sorrow
Izard [Iza91]	Physiological	anger, contempt, disgust, distress, fear, guilt, interest, joy, shame, surprise
Lazarus [Laz91]	Cognitive	anger, anxiety, guilt, shame, sadness, envy, jealousy, disgust, joy, pride, love, relief, hope, compassion
Tomkins [Tom63]	Physiological	anger, interest, contempt, disgust, distress, fear, joy, shame, surprise

The underlying assumption in discrete approaches is that there is a limited set of basic emotions. There are, however, disagreements on the number and nature of basic emotions. Examples of such approaches are catalogued in Table 3.1.

OCC Model

One of the most popular approaches in defining discrete emotions is that of Ortony, Clore and Collins: the OCC model [OCC90]. This model is based on cognitive theories of emotions. Specifically, emotions are taken to represent valence reactions to agents (i.e., whether they are pleased or not about the consequences of an event), events (i.e., the endorsement or rejection of actions) and affinity towards objects (i.e., whether one likes or dislikes an object) [CdRN09]. Based on this, the OCC model specifies 22 emotion types (*joy, distress, happy-for, sorry-for, resentment, gloating, hope, fear, satisfaction, fears-confirmed, relief, disappointment, shock, surprise, pride, shame, admiration, reproach, gratification, remorse, gratitude* and *anger*) and two cognitive states (*love* and *hate*).

The structure of the OCC model is shown in Figure 3.2. This model of emotion is employed in this thesis since it is well-accepted by the cognitive psychological community [Sha08, CdRN09]. The utility of the OCC model lies in its consideration of the different contexts in which an emotion manifests and the way it comes to manifest. It considers the event, agents and objects involved, and characterises the place of the subject therein through accommodating the subject's orientation towards entities in that structure (and/or the structure itself).

In modern IR, given the pervasiveness of application, the plethora of different contexts of use (e.g., entertainment, academic research, shopping, etc.), it is appropriate to use the terminology of the OCC model. This is because it contains (potentially) complex scenarios involving different events, agents, objects, related in nontrivial ways. As discussed in Section 2.4, behaviour/cognition in all such complex scenarios is mediated by emotion, and so for the purpose of emotion-based IR, given the need to understand the context in which emotions arise, the OCC model appears to fit well with the purpose of this thesis.

One of the characteristics of IR is its mediation of social activity, which is inherent in many of the contexts related to above (e.g. entertainment). Even in the OCC model, a model derived from the cognitive experience of the individual, we see that one cannot ignore a subject's orientation with the other people, i.e., emotions are: (1) socially motivated (see Section 3.5); and (2) motivate/influence social activity.

3.4 Emotion and Individual Behaviour

Section 2.4 discussed the current models that take the role of emotion in the IR&S behaviour into account. We described that emotion is considered as a secondary factor in IR&S behaviour, e.g., in Kuhlthau [Kuh93] and Nahl [Nah05] which are the most important models. In order to conceptualise the role of emotion in IR&S, it is important to understand the individual and social nature of the contexts in which emotions arise and the way they come to manifest in these contexts. In this section, we argue that emotion affects every aspect of human behaviour, including rationality, decision-making, attention, and memory access.

This is important since people's information seeking behaviour is mostly assumed to be rational in IR. However, there seem to be several rationalities at hand, depending on the type of information

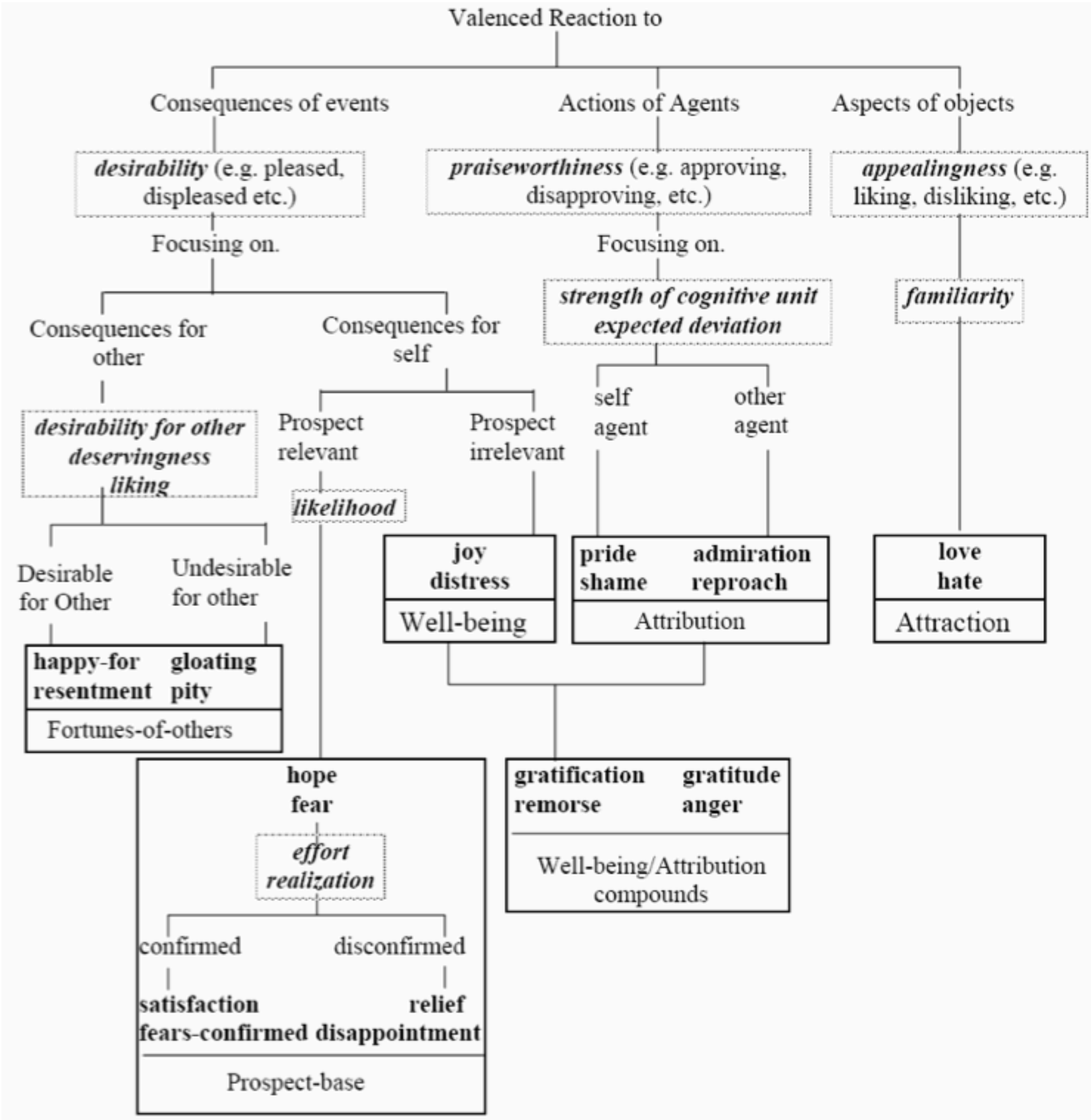


Figure 3.2: OCC model reproduced and provided by Masum [Sha08] [(p. 67)] for ease of reference, considering two original models [OCC90] [p. 19 and 69]. The bold-italic phrases indicate cognitive variables and bold phrases indicate emotion types.

that is sought. For example, behaviour related to entertainment differs from that of academic searching.

In this section, we study the possible role of emotion by looking at the rationality of emotion and its effect on decision-making, attention, and memory retrieval, which are fundamental processes involved in any IR&S behaviour. In particular, we first discuss the rationality of emotion in Section 3.4.1, followed by the role of emotion in respectively decision-making, attention and memory access in Sections 3.4.2, 3.4.3, and 3.4.4 respectively.

3.4.1 The Rationality of Emotion

Interaction in IR is mostly assumed to be rational even if that depends on the type of information sought. On the other hand, emotions seem to be related to a lack of rationality:

> "...[emotions, as feelings,] disrupt our thinking and lead us astray in our purposes. This is what we call the Myth of the Passions: the emotions as irrational forces beyond our control, disruptive and stupid, unthinking and counterproductive, against our 'better interests', and often ridiculous." [Sol77]

The problem is that there are many different perspectives on rationality, so whether emotions are rational depends on the sense of *rational* we choose. Current studies show that emotions play an important role in the psychological life of rational creatures, influencing rational decisions [Els04]. The question here, however, is whether they can themselves *be seen* as rational. In order to make such an assessment, we have to understand what causes to emotion (or affective) reaction. We can look backward for a justification or look forward to deduce their importance in future decisions/actions. These two views of backward-looking and forward-looking analysis are discussed below.

Forward-looking View

Forward-looking rationality is a matter of the instrumental or strategic importance that emotions might have, as put forward by Solomon:

> “Emotions have a purpose in the sense that actions have a purpose – to get something done. . . . The purpose of an emotion, like the purpose of an action, is a multileveled affair. There are any number of goals in an emotion, from the very specific goals in the emotion (‘wanting to see that bastard punished’), to the very general goals of this emotion or the emotions in general.” (p. 36) [Sol77]

Examples of the general purpose of emotion which Solomon talks about includes the fact that fear facilitates a response to danger, anger to insults or wrongs, disgust to contamination, jealousy to infidelity, sorrow to loss. The rationality of the emotions may be assessed from the forward-looking view in terms of how well they enable us to achieve these particular goals. However emotions may have more subtle goals or aims, and may indeed be rational from the forward-looking view in a way that is independent from what seems to be its general purpose.

Emotions can display *strategic rationality*; irrational emotions such as anger might be more effective in securing long-term benefits. For example, consider the desire for revenge which stems from anger. From a short-term self-interested standpoint, a desire for revenge seems irrational. Revenge won’t bring some valued object back, and has high costs, but the urge for vengeance might have strategic value in preventing long-term costs, e.g., if *“I never take revenge or never retaliate, I am an easy target.”* Emotions that support trust, fidelity and promise-keeping put us at a long-term advantage in the face of short-term gains. In this case, they can count as rational from the forward-looking view since they conflict with rational judgement about beliefs.

The strategic rationality of emotion is particularly important in this thesis, since it defines a possible emotional motivation for a searcher to engage in an IR&S behaviour. Knobloch studied such a motivation factors in an experimental setting and the results show that people engage in IR&S behaviour to maintain their negative emotion for strategical reasons [Kno03], e.g., maintaining their anger for near-future argument. From a forward-looking point of view, emotion can be considered as a primary factor that explains why searchers engage in an IR&S process. We further discuss this in Section 4.2.2 to explain the comprehensiveness of what we call an *emotion need*.

Backward-looking View

In backward-looking, the appropriateness of emotions is selected to their eliciting situation. Rationality in this sense is a matter of how well emotions 'fit' their objects. This is a matter of (a) accuracy and/or (b) justificatory status. The backward-looking rationality approach to emotion can give a different reason from that of forward-looking rationality approach.

In an IR context, the backward-looking rationality of emotion explains whether the emotion experienced at different moments of the IR&S process is rational or not. For example, Kuhlthau explains that during the initiation phase there is a burst of negative emotion. The experienced emotion is rational in the backward-looking view since it can be justified by the uncertainty principle [Kuh93]. Therefore, the backward-looking view explains the secondary factor of emotion in an IR&S behaviour (Section 2.4).

From a backward-looking point of view, two types of rationality exist: (1) a reflective rationality which encompasses conscious judgements or beliefs; and (2) a rational intuition which encompasses pre-reflective experience of consciousness [Sol77]. Therefore, intuitions – *"gut feelings"* – serve as an adequate *rational* justification for the current emotions even without being accompanied by the reflective endorsement in the form of a conscious judgement. This suggests a *two tier* system of evaluation: (1) evaluation involving emotional intuitions; and (2) evaluation involving conscious evaluative judgements.

The system of the first type is valuable in alerting us quickly and at little cognitive cost, to things of potential importance in our environment. However, since these emotional responses are relatively blurred, there is a natural need for a more discriminating evaluative system, i.e. the second type of system. In an IR context, the primary evaluation of selecting documents in a given list can be seen as case (1), since the information presented to the searcher is not sufficient to make an informed evaluation.

In summary , we described how backward-looking and forward-looking views of rationality of emotion can explain the primary and secondary role of emotion in IR&S behaviours. In the next section, the role of emotion in decision-making is studied.

3.4.2 Emotion and Decision-Making

There exists an important place for emotion in decision-making, even though the logical view of rational decision-making view holds no place for it. Again, following to Solomon:

> "formal logic will, by itself, get us to the best available solution to any problem. An important aspect of the rationalist conception is that to obtain the best results, emotions must be kept out. Rational processing must be unencumbered by passion." (p. 171) [Sol77]

However, the varieties of options and outcomes available is typically vast, and we therefore need a non-rational way to delimit the range of things that are to be considered and evaluated, which bring us from thinking to acting. For some philosophers, such as de Sousa, this is precisely the role of emotions: *"in the process of rational deliberation itself, emotions render salient only a tiny proportion of the available alternatives and of the conceivably relevant facts."* (p. 276) [Sou90]

Emotions play this role because they are 'somatic markers': visceral and non-visceral sensations that are attached to images and ideas, and which limit options and direct reflection [Dam94]. This is supported by empirical evidence of the inability to make decisions in patients with frontal lobe damage.

Every cognitive action made by searchers in an IR&S process is (consciously or subconsciously) based on a decision. For example, in an ad-hoc IR scenario, searchers decide to engage in an IR process, decide on the queries they submit, decide on how far they want to browse through the search result lists, decide on which document from the list they want to explore, decide on the relevance of the explored document, decide on whether to continue the search process or not and so on.

3.4.3 Emotion and Attention

There are things that *happen* to us, rather than things we do. Emotions alert us to important or significant things by capturing and directing our attention. This is an automatic and reflexive response to stimuli. However, emotions do not just reflexively direct and focus attention; they also allow us to keep attention focused on objects, and make it very difficult for us to shift our focus. Neurological evidence for this is plentiful, and points to the fact that emotions are associated with an increase of cortical arousal[6]:

> "emotional reactions are typically accompanied by intense cortical arousal ...this high level of arousal is, in part, the explanation for why it is hard to concentrate on other things and work efficiently when you are in an emotional state (p. 289) ...arousal locks you into whatever emotional state you are in when arousal occurs" (p. 290)[LeD96]

One result of this arousal is the increased processing of information we receive from emotional stimuli. The point of keeping attention is not to detect significant objects and events, but to promote a more detailed analysis and evaluation of the object or event already identified. For example, in an IR context, in the case of relevance assessment in an IR&S process, searchers' emotion plays an important role. This argument was supported by the results of experiments conducted by Arapakis et al. [AKJ09] where they show that the emotion captured from searcher while performing an IR&S process can act as a relevance feedback.

3.4.4 Emotion and Memory Retrieval

Affects, particularly moods, are argued to mediate in human information processing, particularly memory retrieval. Several models have been proposed, such as the *associative network* model [Bow81, Ise84], and the *feeling as information* model [SC88]. we briefly discuss these below.

Associative Network Model In the associative network model, moods are cues for memory retrieval, i.e. they activate other pieces of memory where there is information associated with a similar mood. In addition, people in a positive mood are more likely to remember positive concepts and events than when they are in a negative mood [SS88][7]. This is particularly relevant

[6]Increased arousal is (in part) a matter of increased sensitivity of cells in the cortical and thalamic regions, and this can result in a 'feedback loop': sensitivity triggers further arousal of the amygdala.

[7]Despite the fact that we index our memories by feelings, memory itself has more of a graph structure

for the subject of this thesis, since it indicates that the current emotional state of searchers can influence their memory and hence their needs and actions.

Feeling as Information Model The feeling as information model assumes that people use their feelings as a source of information when making judgements. In this model, situations in which moods should be considered as information have been discussed. For example, when moods are irrelevant to the task or made less salient, they lose their information value [GAT98].

The mediation of human information processing by emotion is especially relevant for us since the purpose of IR systems is to support this information processing[8]. Thus, if moods are used for memory retrieval, as stated in the associative model, we can say by extension that they are also seen in the computational information behaviour of users.

3.5 Emotion and Social Behaviour

Emotion is one of the essential aspects of the communication process since it has a significant role in virtually every aspect of the human social life. In this thesis, we argue that emotion is a primary factor in the IR&S behaviour, i.e. it is the reason why searchers engage in an IR&S processes. This argument is explained in particular with evidence from sociology. We explain our argument from two viewpoints: (1) human beings are able to communicate emotion (in Section 3.5.1), understand and feel the communicated emotion (in Section 3.5.2); (2) the emotion state of a human being can motivate them to engage in different social actions including seeking information and/or emotion from other people (in Section 3.5.3) or via entertainment-based activity (in Section 3.5.4).

3.5.1 Communicating Emotion

Dillard argues that *"human beings evolved to meet adaptive challenges posed by the environment"* [Dil98], and: (a) that affect primarily guides behaviour and that its evolution corresponded

[8]The cognitive theories of human information processing and IR/computing research have shared a similar concept on this ground.

to improving humans' interaction with their social environment; (b) that this evolution is in the presence (and service) of social interaction.

This *social interaction* when purposeful or consciously directed towards another human being, becomes a communication process. The phenomenon of an agent understanding another through such a process, including, for example, the understanding of the other's emotional state, is referred to as empathising [Dav06]. This phenomena accommodates the human being in coping with unfavourable emotion states, and maintaining favourable states, i.e., mood management; the phenomena of empathy coping and mood-management are described in the following section.

Darwin was one of the first to study emotion, its origin and expressions [Dar56]. This early work was an important step but the ideas have been challenged by many. He believed that emotions are expressed and not communicated, meaning that *"emotions are the actual mover and the expression is just a by-product"* [ED94, LO92]. This point of view was due to his belief that emotions are inherited from animals and serve the process of evolution. Thus, the focus on the *biological continuities between human and animals* and the emphasis on *unintentional displays of emotion* caused him to believe in expression of emotion rather than communication of emotion.

However, recent studies have argued that not only do emotions have intentionality, but also that the intentionality of emotion is crucial. At an abstract level, scientists believe that humans are able to communicate emotions and that this communication happens via facial, vocal, body, physiological, and verbal cues, often though a combination of different cues [PK02]. Communication therefore requires the ability for humans to observe these cues (for the purpose of understanding), and this is known as empathy.

3.5.2 Empathy

Empathy, similar to relevance, is a very complex human concept. It is a *"phenomenon in which one individual, through observation of another, comes to experience some change in his or her thoughts or feelings"*. ([Dav06], p. 443)

There are three groups of approaches for explaining empathy. The first group considers empathy as an *essentially emotional phenomenon*, meaning that observation of another person, results in experiencing a changed emotional state in the observer (e.g., [Bat91]), which in some cases is similar to the emotional state of the target (e.g., [ES87, Dav06]).

The second group considers empathy as an *essentially cognitive phenomenon*, meaning that observation of another person results in understanding the target's internal state without necessarily experiencing any emotional state in the observer (e.g., [Wis86]).

The final group considers empathy as a *multidimensional phenomenon "including both cognitive and emotion components"* (e.g., [Dav83, Dav94, Hof84]). According to this point of view, the observation of another human being can result in some response on the part of the observer which could be either cognitive, affective, motivational, or behavioural [Dav06].

Empathy is one of the most fundamental capabilities of the social behaviour of human beings, since without it the communication of emotion would not be possible. In context of IR, the understanding of the emotion expressed in a document (e.g., review of a product) by a searcher is due to this capability.

3.5.3 Coping

Coping is a process in which an individual tries to alleviate or diminish an undesirable psychological (particularly emotional) state such as stress ([LL94], p. 152). Lazarus defined coping as *cognitive and behavioural efforts to manage specific external or internal demands (and conflicts between them) that are appraised as taxing or exceeding the resources of the person* ([Laz91], p. 112). Coping strategies are categorised into three main groups: appraisal-focused, problem-focused, and emotion-focused coping strategies [WW11]. We describe bellow these strategies.

Problem-focused Strategies

In problem-focused strategies, the cause of a problem is challenged by individuals. In this way, an individual the negative emotion such as stress or anxiety by gathering information about the problem and employing their newly-learnt skills [CSW89]. Section 2.4 explained that Nahl [Nah05] explicitly consider an IR&S process as a coping process. Since her investigation was in the context of satisfying an IN, i.e., finding information that alleviates the negative feelings associated with a lack of knowledge, we argue that the current IR&S process can be considered as problem-focused coping strategies. This scenario was also described by Kuhlthau, where she suggested that the IR&S process has the purpose of satisfying searchers' IN by reducing the uncertainty and as a result diminishing the negative feelings such as anxiety, and frustration associated with it [Kuh93].

Appraisal-focused Strategies

In appraisal-focused strategies, an individual belief is reshaped by oneself, e.g. by act of denial, or by avoiding from facing problems [BM84]. The way of thinking about a problem may also be changed by modifying one's values and targets, for example by finding the comical aspect of a situation [EP90]. This is coherent with the *cognitive dissonance theory* of Leon Festinger, which had a great influence on social psychology and is one of the most-studied theories in the field [Fes60]. Festinger was interested in finding out what happens when people experience psychological inconsistency. He claims that the inconsistency causes an unpleasant emotional state, a dissonance; and he argues that people act in a way so as to reduce dissonance[9].

One of the common approaches in appraisal-focused coping is to seek others help in an emotionally challenging situation [RCH02]. Rimé et al. [RCH02] explain such social motivation with cognitive dissonance theory and suggest that seekers are searching for either a justification of their dissonance or for a new form of belief about their problem. They show that the more challenging the experience, the more people are willing to share it.

In the context of IR, manifestation of appraisal-focused strategies is apparent. People explain their emotionally challenging experiences in the internet via reviews, tweets, comments in social networking sites (e.g., Facebook), personal blogs, and so on. Others, instead of sharing, may search in such pieces of information to find similar experiences, connected along with their associated responses, so that they can be used in their coping process.

[9] The most famous example of this is his early study in *When Prophecy Fails* about a group of people who believe in an impending fixed the end of the world to its impending end [FRS64] . When that prediction failed, they experienced inconsistency and therefore they took steps to make the contradiction go away by recruiting converts.

Emotion-focused Strategies

Emotion-focused coping efforts are aimed by individuals at altering feeling of their particular person-environment relationship [Laz91]. The strategies used by an individual in emotion-focused include engaging in hedonistic activities, managing negative feelings, and/or using arousal balancing (e.g., relaxation) procedures [WW11]. The idea here is to engage in an activity that stimulates positive and uplifting emotions in the person.

Manifestation of emotion-focused strategies is also apparent in an IR context. People seek for emotionally-rich data such as music, movies, video clips, mainly for their emotional stimuli and IR applications are the main applications used to access to such documents by searchers.

Discussion

Coping is an important concept for this thesis. We argued that current models from IR/IS that explains the role of emotion in IR&S behaviour in satisfying an IN can be categorised as problem-focused strategies, since the current view of the role of the emotion in such behaviour is to diminish the negative feelings created by uncertainty via finding relevant information [Kuh93, Nah05]. However, there are other types of coping strategies which can also motivate searchers to engage in IR&S behaviours. Researchers show that in the digital age, different coping strategies have manifested on the internet [LEG01], e.g., searching for similar experiences in reviews, blogs, tweets, Facebook comments, etc. (appraisal-focused coping strategies) and/or searching for emotionally-rich contents such as movies, music, news, etc. (emotion-focused coping strategies).

3.5.4 Mood Management, and Disposition Theories

Due to the pervasiveness of the emotionally-rich contents on the Internet (e.g., news, music, movies, etc.), it is important to understand the IR&S behaviour which has an entertainment aspect. It is not controversial to state that entertainment has an important role in human life. The consumable media (e.g. television, news, etc.), which is particularly easy to access nowadays, provides people with vast varieties of content, making it the most popular type of entertainment. It has therefore been of particular interest for scientific study to understand its uses and effects on people. Dolf Zillmann was instrumental in its development and the establishing of a range of theories aiming to address two academic questions: (1) what is entertainment, its characteristics, and its effect on its audience?; (2) why are audiences so attracted to it? [Vor03] Both these questions are of relevance to this thesis, in particular the second one, since it allows better understanding of an entertainment-based IR&S behaviour. For this reason, we further inspect some of the most accepted theories in this regard. More precisely, the theory of *mood management*, which explain why and how the audience seeks entertainment; and that of *affective disposition*, which explains how entertainment affects its audience are discussed below.

Mood Management

The underlying speculation of the mood management theory is that to a great extent, human behaviour is dominated by hedonic motivations [Oli03, Zil88b, Zil88a, ZB85]. According to this theory, human beings try as much as possible to maximise or maintain their pleasure and diminish or alleviate their pain by arranging their environment [Oli03]. One form of environment arrangement is a symbolic one, whereby media content is arranged for consumption. Phrases such as *"I am in the mood for a comedy"* are commonly used by people to indicate affinity towards their entertainment preferences. Entertainment preferences influenced by person's taste, habits and long-term interests, are not necessarily static for any individual [Oli03].

One important difference between mood management theory and other related theories is the assumption that individuals are not always aware of their hedonic motivation, and thus cannot articulate it [Zil88b, Zil88a, Zil00b, ZB85]. Having this assumption in mind, and being aware of its consequences on the experimental studies strategies (i.e., employing behavioural measures rather than self-report technique), Zillmann et al. [Zil88b, Zil88a, Zil00b, ZB85] confirmed that people do not organise their media environment randomly. Instead, hedonic motivation is a key factor in affecting the entertainment selection process [Zil88b, Zil88a].

Human States that Need Regulation

We have explained earlier that mood management theory posits that entertainment choices are a reflection of a basic human need to enhance or retain positive states, and to lessen or steer clear of negative ones [Oli03, Zil88b, Zil88a, ZB85]. Thus researchers try to understand and define positive and negative states of users that may play an important role in media selection, and understand and define the characteristics of the media which may be selected to regulate these states [Zil88b, Zil88a, ZB85]. They have suggested two possible states in which there may be a need for regulation: physiological arousal and affect.

In the case of physiological arousal, mood management theory suggests that users might be over-stimulated (i.e., stress) or under-stimulated (i.e., boredom). An individual experiencing such states will choose their entertainment content according to their expectations of what would lead them back to an optimal state [Oli03].

In the case of affect states, mood management theory suggests that users might be in negative (i.e. dysphoric) or positive (i.e. upbeat) moods. An individual experiencing negative affects will choose entertainment content that helps them to alleviate or diminish the negative mood and those experiencing positive affects will choose entertainment content that helps them to intensify or prolong their state [Oli03].

For the case of media content, mood management theory considers numerous characteristics: (1) entertainment excitatory potential (2) its hedonic valence (3) its semantic affinity to the viewer's current state, and finally (4) its absorption potential [Zil88a]. The entertainment excitatory potential is the most relevant characteristic for arousal regulation whereas the other three are the most suited for mood regulation [Oli03]. Mood management theory dictates that individuals who are overstimulated tend to choose calming media content while those who are under-stimulated tend to choose exciting content [Oli03].

In the case of affective states, individuals who experience a negative affect, for example, tend to choose positive or uplifting valence media contents as a way to increase their level of gratification. In addition, the chosen media contents tend also to be semantically far from the circumstances that created the negative affective state. On the other hand, in the scenario of experiencing positive mood, the theory suggests that maintaining a positive affective state with media contents that have positive valence tones and high levels of semantic affinity will be the optimal set of options [Zil88b, Zil88a].

Two important conclusions are drawn from this evidence: (1) emotion acts as a primary factor in IR&S process, i.e., as an autonomous need (see Section 4.2); and (2) the emotion aspect of documents is an important factor in satisfying searchers' needs, particularly emotion need.

The second conclusion motivated this thesis to consider the emotion aspect of documents as a new feature to determine whether a document would satisfy a researcher's need (see Section 4.3.3).

The above has clear implications in IR. For example, an IR system can improve its effectiveness by analysing the previous history of a searcher with respect to the emotion and semantic aspect of the visited documents and based on that, predict which documents should be recommended to the searcher. This thesis conducts such experiments and the results support this example (see Chapter 6 and 7).

Disposition Theory

Zillmann et al. defined an additional media content characteristic which has a great impact on individual gratification. They found that the narrative element of the media content is in fact a crucial characteristic that causes an emotional experience in the audience. The comprising factors in viewers' experience of media entertainment are: (1) the disposition of the viewers in relation to the portrayal of characters; and (2) the outcomes which the characters experience [Zil85, Zil91, Zil00a, ZC77]. They proposed a *disposition theory*, which explains the possible gratification level based on the outcome of the story. Based on this theory, *"when liked or beloved characters are shown experiencing positive outcome or reward, or when disliked or hated characters are shown experiencing disappointment or negative outcomes"* ([Oli03], p. 88) the viewer will experience the highest level of gratification, whereas *"when liked characters are shown suffering and when disliked characters are shown prevailing"* ([Oli03], p. 88) the viewer will experience the lowest level of gratification [Oli03].

Problematic Media Genre for Mood Management Theory and Possible Reasons

Numerous studies have shown a great support for the underlying assumption of mood management theory [Oli03]. It also intuitively makes sense, as the purpose of entertainment is to have a *"good time"*. However, a variety of genres and media contents deviated considerably from what would be normally considered pleasant or upbeat, at least superficially. In particular, there is popular media content with genera such as *Crime Dramas*, *Violence and horror*, or *Tearjerkers in Movies and Music* that is problematic for mood management theory. These genres portray suffering, pain, fear, terror, and sadness, and could stimulate negative emotions in their audience. Further taking their popularity into account, this conflicts with the basic speculation of mood management theory. There should be either a level of experiencing gratification in the viewers of this type of media content, or other explanations should exist which may contrast with the assumption of hedonism mood management theory.

Several studies have investigated possible reasons for audience interest in so-called negative stimuli genre media contents. Boyanowsky et al. [BNW74] suggest that this genre media content provides a sage context that allows people to cope with their fear or anxiety. Finigstein and Heyduk [FH85] suggest that in some situations violent programming may be consoling since it justifies hostility and aggressiveness. Among these problematic genres, *Tearjerkers in Movies and Music* appears to be the most inconsistent with mood management theory. Oliver [Oli93] suggested that the greater the emotional responsiveness to others' suffering, the greater the feeling of gratification experienced by this sort of entertainment. Evidently the saddening experiences may gratify some viewers, largely those who have understood and shared the feelings of another in the past [Oli93].

Individuals do not always follow the hedonistic assumption underlying mood management theory, i.e., boosting their pleasure and avoiding the pain causing stimuli, e.g., in situations such as a funeral, the negative affect is the most accepted and appropriate response [EE00].

The *mood input* model, suggested by Martin et al. [MD98], explains that individuals are seeking the positive outcomes more than positive moods themselves, and such positive outcomes may be indicated by negative emotional experiences under some difficult circumstances. For example, in a situation where an individual's desired outcome is to feel compassion, the experience of feeling sorrow in response to another's loss may be regarded as a sign of the achievement of the desired outcome [MD98]. The proposal made by Mares et al. [MC92] suggests that there is an increase in self-esteem as a result of one's self-comparison to those less fortunate. This statement is in line with Festinger's (1954) social comparison theory [Fes54].

Zillmann'a Explanations for Problematic Media Genre

Explanations provided by Zillmann in support of his mood management theory for user behaviour in selecting problematic genres are as follows: (1) due to the disposition theory, the viewers of such genres may experience the highest level of gratification, as a result of the succession of beloved or liked characters in getting justice, as in crime dramas, or in dealing with a problem or difficulty or their survival, as in violence and horror genre [Zil00b, ZW85]. Thus, experience of the negative feeling of anxiety and suspense throughout the movie may in fact work in support of intensifying the utmost level of gratification [Zil80, Zil91, Zil98, ZHB75]. (2) there is a distinction between those individuals who seek for immediate gratification and those who extend the duration of their negative state for some future positive outcome. (3) some counter-hedonistic choices are based on individuals' need to seek information rather than be exposed emotionally to the media entertainment [MC92].

In summary, Zillmann et al. suggest that if individuals attempt to *repair* their mood, the hedonic concern, is the key driving force for the selection of most entertainment material [Zil00b]. Entertainment theory revolutionised the way of thinking about human needs and desires. Questions about how and when mood management occurs not only provide an insight into viewers' entertainment choice, but also examine fundamental questions concerning human need and desires.

The theories, evidence, and discussions presented in this section justify the role of emotion as a primary factor in entertainment-based IR&S processes, since it provides the hedonistic reason behind such processes. Therefore, it is important to consider emotion as an individual and autonomous need for searchers to engage in an IR&S process. In the current views in IR/IS presented in Chapter 2 the primary of emotion has been fairly ignored. In Chapter 4, we define the concept of *emotion need*, which encompasses the primary factors of emotion mentioned in this chapter.

3.6 Chapter Summary

In this chapter, we explained the role of emotion in the personal and social life of human beings. We particularly looked at the nature of emotion, followed by its conceptualisation models. Its role in fundamental aspects of human behaviour was then discussed (rationality, decision-making, attention, and memory retrieval). Next, the importance of emotion in social life was explained (communication of emotion and concepts such as empathy and coping). Finally, we investigated the role of entertainment, focusing on media content in the coping process.

The study showed that the current view of the role of emotion in IR&S behaviour can be supported by one of the strategies of coping, i.e., problem-focused strategies. However, the pervasiveness of the use of IR applications for the purpose of entertainment and the existence of emotionally-rich data on the web provides evidence that some information seeking behaviour can be categorised under other strategies of coping, such as arousal-focused and emotion-focused. Considering these strategies in an IR&S behaviour process can lead to better satisfaction of the searchers' needs.

Part II

Theoretical Contribution

The information seeking behaviour, the nature of emotion, its conceptualisation and its importance in personal and social life have been discussed in Part I. In this part which contains the theoretical contribution of this thesis, a new relevance concept is proposed to encompass (1) emotion-based phenomena search context and (2) emotion-based techniques for effective retrieval. This is important since the effectiveness of information retrieval and seeking systems can be improved through the use of a better understanding of searchers' needs (such as emotion need) and better document representation (such as emotion representation).

Chapter 4

Emotional Relevance

4.1 Introduction

In this section, we merge the notion of emotion into the existing IR framework, by making it central in the system of needs of a searcher, and therefore as a factor that ought to be considered by the IR algorithms. We discuss a realistic system of user needs by incorporating not only theories from IR/IS domain, but also from research in psychology, philosophy, sociology.

In Section 3.5 we explained that emotion aspect of documents plays an important role in satisfying searchers' needs. For this purpose, It is important to consider this aspect of documents in IR. We argue that the emotions extracted from documents can be used as an emotion representation of the documents. Further, considering this aspect of documents can act as a clue to a better understand and answer of searchers' needs.

This chapter explains the *searcher's need system*, and the concept of searcher's *emotion need*, the *emotion object*, and the *emotional relevance*. It then presents a conceptual map for the proposed emotion relevance, explaining the searcher and system side in different prototypical scenarios.

4.2 Searcher's System of Needs

Following Wilson [Wil93] (Section 2.4), we argue that three types of human needs constitute the searchers system of needs: physiological need, informational need and emotion need. We define them as follows.

Definition 4.1 (Physiological Need)

The low level physical, animalistic requirements for survival, such as hunger, trust, avoidance, sexual desire, and the elimination of bodily waste.

Definition 4.2 (Informational need)

An individual or group desire to acquire information[1] for the purpose of addressing a lack of knowledge.

Definition 4.3 (Emotion State)

An emotion state is a set of emotions that a human being experience at a given point in time.

Definition 4.4 (Emotion Need)

An individual or group's desire to be in a particular emotion state by means of acquiring information, *and/or emotion*[2].

Among these needs, emotion need takes a central position since it underlies other needs, or other needs depend upon it in some ways.

4.2.1 Centrality of Emotion Need in IR&S Behaviour

The existence of other needs than the IN, such as affective need[3], in the human need system, and its interrelation with information need have been discussed [Wil93] (see Section 2.4.3). Although other works tend to not explicitly refer to affective need, they discuss the role of human emotion in an information seeking process, e.g., Kuhlthau's model [Kuh93]. Current research: (1) observes that participants experience a burst of negative feelings due to uncertainty associated with vague

[1] we use the definition of information proposed by Mizzaro [Miz96] (Section 2.3).

[2] We motivate this ability of emotion to be individually acquired in the following

[3] *"affective needs (sometimes called psychological or emotional needs) such as the need for attainment, for domination etc.;"* [Wil93] centres on the *emotion need* defined in this thesis.

thoughts leading them to recognise that they have an information need; and (2) agrees that there is a positive correlation of successful information seeking processes with the decrease of these negative feelings [Kuh93]. The coping strategies discussed in Section 3.5.3 capture such behavioural patterns, but are more general. The IR/IS behaviour discussions, with respect to the relation between affect and cognition, can be categorised as being among the coping strategies [Nah05]: more precisely, they are part of the *problem-focused* strategies (see Section 3.5.3).

As we discussed in Section 3.5.3, there are other strategies, namely *emotion-focused* and *appraisal-focused*. Given the pervasiveness of social applications on the web (such as Facebook and Twitter) and the diverse range of documents (movies, music, images, news, blogs, etc), the infrastructure exists on the web for the accommodation of these other strategies. In modern times, computers are used to fulfil daily social needs: entertainment, dating, getting to know people, maintaining friendship, gaming, etc. Researchers show that people do actually employ the coping strategies mentioned above, in their daily technology-infused lifestyle [LEG01].

This ever-increasing consumption of entertainment in the web, and its link to the discussion in Section 3.5.4 on mood management, suggests that searchers, consciously or subconsciously, work to satisfy other needs than informational ones, particularly to maintain their emotional state [Oli03]. Also, in Section 3.5.4, we explained that due to strategic reasons, humans exhibit tendencies to want to experience particular emotions (usually negative), e.g., keeping anger for an upcoming conflict or maintaining sadness in a memorial ceremony.

These IR&S behaviours brought us to rethink and acknowledge the existence of the emotion need as a central need in IR/IS, and about what a proper definition should capture. An adequate definition of emotion need ought to consider sociology of emotion. Our current definition of emotion need takes this into account.

4.2.2 The Comprehensiveness of the Emotion Need Definition

The definition of emotion need is discussed from two perspectives: (1) current theories of emotion in information retrieval and science; and (2) current theories of emotion in sociology and communication. This definition covers the information retrieval and science domains. As discussed earlier, the role of emotion in the information seeking process is to alleviate and/or diminish the negative feelings experienced because of uncertainty, so the emotion need here is for experience positive feelings of satisfaction via obtaining information.

In the case of general mood management and disposition theories, the emotion need of the user is to maximise the experience of positive emotions and minimise the experience of negative emotions. The emotion need here is different to the one in IR/IS theories, the difference being that the emotion part of the documents in this case is a primary factor for searchers (See Section 3.5.3 and 3.5.4). Our emotion need is not limited to positive emotions only (as we explained earlier), as searchers may want to be in negative emotion states (i.e. in the case of strategic emotion).

The searcher's emotion need could be to maintain the current emotion state or change it. The emotion need of a searcher could be similar or different from what the searchers experience at the start of IR/IS processes; e.g., in Kuhlthau's model [Kuh93], there is doubt, anxiety and frustration, whereas their emotion need is that of joy, happiness and satisfaction. On the other hand, in the case when searchers are feeling *down*, the emotion experienced is sadness and the emotion need can be to maintain this sadness [Oli03]. For example, in situations such as funerals, when negative affects are the most suitable under the circumstances, individuals may seek for stimuli for maintaing their sadness [EE00].

4.2.3 The Relationships Between Needs

The relationship between physiological and emotion need is thoroughly addressed in the psychology literature [Iza91]. In contrast to Wilson [Wil93], we argue that physiological needs are not directly satisfied through an information seeking process, but that they instead lead either to an emotion or information need that initiates the information seeking behaviour which goes on to satisfy these needs. For example, hunger (i.e., physiological need) can lead to either searching for close-by restaurants (i.e., information need) or negative emotion states (e.g., frustration) needing to be resolved by watching funny clips (i.e., emotion need). Due to this delegation of physiological need to information or emotion need, we do not further investigate physiological need. Therefore, all we need is to investigate the relationship between information need and emotion need.

We argue that an emotion need is more fundamental than an information need in the sense that if an information need exists it implies that there is an underlying emotion need to satisfy this information need. The whole IR&S behaviour is thus driven by an emotion need. However, the converse may not necessary be true, e.g., a user could want to be happy/sad/angry but without having a well-defined *IN*. Thus, whenever information need is discussed, an emotion need is pre-existent.

In the case when the emotion need of the searcher is to diminish the negative feelings associated with a lack of knowledge (i.e., an IN), the emotion need would be satisfied if the IN associated with it is resolved. For example, if a searcher's IN is to know about topic x, the searcher must *believe*[4] that information about x has been acquired, in order for their emotion need to be satisfied. Thus, the emotion need will not be resolved unless the underlying information need is resolved, since in this context, the information need is the dominant one.

4.2.4 The Emotion Need Spectrum

There are in fact emotion needs that do not imply an information need in the way we have defined information above. An example of such needs are the scenarios explained in Section 3.5.4, i.e., users who are stressed and look at some clips that they know will help to relieve their stress. Of course, one way of remembering these clips is by employing the associative nature of the relationship between emotion and memory (see Section 3.4.4). Other ways include looking at the popular (most viewed/highly recommended) objects. In all these scenarios there is no particular information need to be resolved, but only an emotion need, e.g., when searchers are seeking for funny clips in YouTube. In these scenarios, it is argued that the emotion aspect of information objects is more important than their information aspect, and we label them as *extreme emotion need* scenarios. Thus, one can present emotion need as a continuous spectrum ranging from informational needs to extreme emotional needs.

It is important to differentiate between extreme information need and extreme emotion need as two motives for initiating an information seeking. As discussed above, the emotion aspect of the object is what is important for the extreme emotion need scenarios, in contrast to the extreme information need where it is the intended information aspect that is important.

It has been shown that information need motivates searchers to engage with an IR system. An emo-

[4] see Section 3.2.2

tion need can be a motive for searchers to use an IR system when it manifests itself as information need. It is our belief that emotion needs, even when they do not lead to a particular information need, can motivate searchers to use an IR system.

4.3 Information Object or "Emotion Object"

As we know that emotion states can be changed or maintained by emotion and/or cognition, a piece of information or a particular emotion can potentially cause these changes. Documents consisting of both information (e.g., topics) and emotion can therefore potentially cause such changes. The information representation of documents (i.e., information object) has been extensively studied in IR. In the next section, we discuss whether, analogically, documents can be taken to consist of emotions and thereby represented in terms of them (i.e., emotion object).

In order to create an emotion representation of a document, the emotion aspects of it need to be extracted/analysed. The possibility of extracting emotion from content of a document is the main practical challenge. This challenge is explained in Section 4.5 and the accuracy of the emotion extraction system is investigated in Chapter 5.

In this section, we assume that it is possible to extract emotion from documents based on their contents, and that the extracted emotions represent the document itself. We justify our argument by discussing two other possible arguments, i.e., extracted emotion: (1) represents the emotion of the creator of the document while making it; or (2) represents the emotion of the observer of the document when interacting with it. We further justify our argument based on the cultural dependency of emotion, and finally, we define *emotion object* and *emotion relevance* concepts.

4.3.1 Document Emotion a Representation of Creator's Emotion

It can be argued that the emotion extracted from a document represents the emotion of its creator, e.g., for the opinion-based texts such as blogs, reviews (e.g., a movie or product review) and tweets. Our argument is that the extracted emotion represents the message that authors of such documents want to transmit rather than the authors' emotion while writing them. Therefore, it is not correct to associate extracted emotion to the creator of an opinion-based document.

This argument is also valid for documents that are not directly considered to be opinion but still contains emotion, e.g., news, movies, books, etc. For example, a news article explaining a natural disaster such as a tsunami and its consequences (killing thousands of people) clearly has associated emotions. However, it is difficult to relate the emotion extracted from such documents to a journalist reporting such news. This is because the journalists reporting news convey emotion in their articles that could be different from their experienced emotion while writing the article. Therefore, there is not necessary a similarity between the emotion expressed in the text of the news and in the mind of the journalist at the time of writing/presenting it.

There are two reasons underlying the *possible difference*[5] between the emotion transmitted/conveyed in a document and the emotion experienced by the creator while making the document: (1) people often may not express their true feelings [Sul02], therefore, they express emotion differently to what they really experience; and/or (2) the emotion experienced by the creator while making the documents sometimes is not important or relevant to the message that he or she wants to transmit. For example, a journalist who is under pressure to finish an article (i.e., experiencing stress), does not necessarily experience an emotion similar to the emotion of the message about an improvement in a new health care system which he tries to report in his article. In this case, his emotion can be considered irrelevant to the message.

4.3.2 Document Emotion as a Representation of Observer's Emotion

It can be argued that the emotion extracted from a document represents the observers' emotion, thereby indicating that the document's emotion is highly objective, i.e., assuming that all observers experience emotion similar to the extracted emotion from the document. This argument is obviously wrong. Emotion features captured, for example, via monitoring readers' biometric features and/or facial expressions show the subjectivity of the emotion experiences. Arapakis et al. [AAJ10] showed that the captured emotions from such features are subjective and can indicate personal preferences. Observers can not only: (1) experience emotion more independently from

[5] We do not argue that the emotion extracted from the document never matches the emotion of its creator while making the document, but we argue that there could be situations where these two emotions do not match.

those extracted from documents; but can also (2) experience different emotions from one another.

We argue that the emotion experienced by observers is neither objective nor completely subjective, but it is intra-subjective. If emotion was completely subjective, then it would be impossible, for example, for the writers to write something since they could not anticipate what the message the readers would get after reading their document. What makes such communication possible is the knowledge of the creator of a document about his/her audiences.

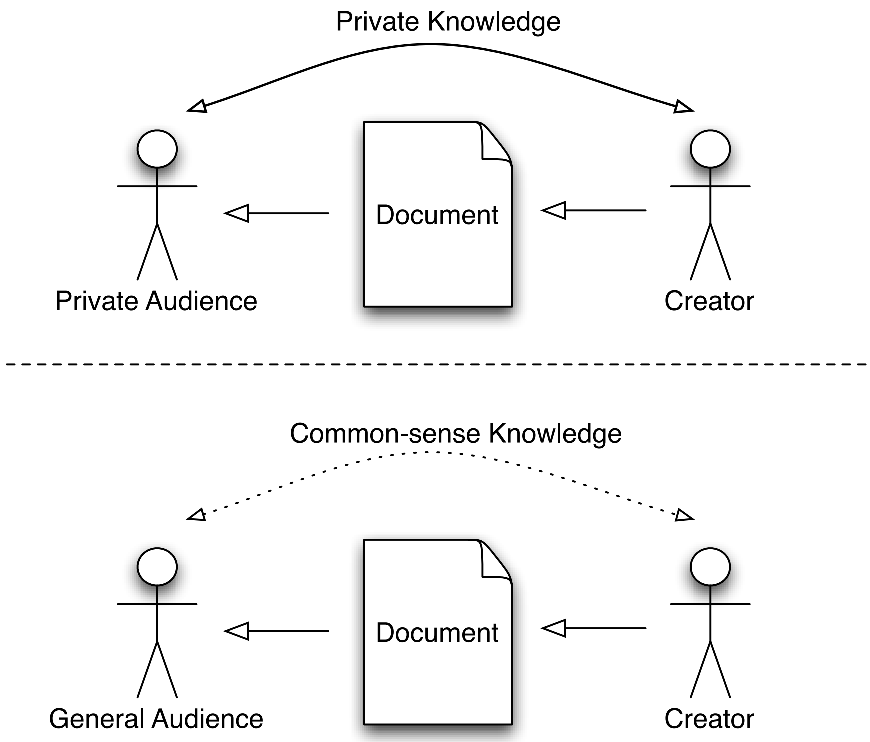


Figure 4.1: Creator-Audience Relationship: The knowledge used by the creator to communicate with the audience

To understand the last point we now focus on the act of creating a document, i.e. the act of writing. In the process of authoring, there are two possible types of audiences an author can write for, namely a private audience or a general audience. Private audiences are the ones with whom the author shares private knowledge. For example, they both agree on the fact that the word dog represents a funny or scary concept. General audiences are the ones with whom the author shares a common cultural context, i.e., a *common-sense knowledge* (cultural knowledge). They have to follow general conventions of society. This convention is referred to as common knowledge or common-sense. To this category belong the majority of objects: movies, music, and many textual documents whose audiences are general public, e.g., books, news, etc.

Common-sense represents the norms of the belief of a society with respect to concepts, actions, named entities, etc. For example, stealing is bad, helping is good. This common knowledge is the glue that allows observers to understand the authors. For instance, according to common-sense, a tsunami is a natural disaster that is potentially destructive: it is a sad and unpleasant event. The closer one's belief system is to the norm of the society, the more similar their experienced emotion to the ones based on common-sense is. An observer's emotional experience, even if is different, is complementary rather than contradictory information.

4.3.3 Document Emotion as a Representation of Document Itself

In the preceding section, we have shown that the emotions expressed in a document may be different from the emotions the creator experienced while making it, and from the emotion of the observer at the time of consumption. We argue that a third view exists, in which emotions can be assigned to a document and considered as attributes of that document.

Sections 3.5.3 and 3.5.4 explained that a piece of information or a particular emotion can potentially maintain or change an emotion state. Given that a document can carry both information and emotion, it can be argued that the emotion part of a document is what can satisfy an emotion need, especially in the case where there is no specific information need in mind. For example, if one has an emotion need to *"get happy"* and one is looking for *"funny clips"*, there may be no specific information to look for, and instead the emotions associated with the documents are what are relevant to the satisfaction of that need. Thus, there is a requirement to expand the traditional concept of relevance, and we do so by introducing the notion of emotion relevance.

Definition 4.5 (Emotion Relevance)

A relation between the emotion aspect of a document and the emotion need of an individual

Associating Emotion to Parts of a Document

Emotion, in a way similar to topicality, can be associated to parts of a document (a subset of physical features). The nature of the features of these parts depends on the type of document (e.g., text or multimedia) and the model used to extract emotion. For example, in a movie, an emotion object could be the continuous subset of physical features of the movie such as sound, colour, images, or the dialog text of a scene of a movie from which can be associated with a set of emotion features. In a textual document an emotion object could be a word, sentence, phrase, paragraph or the whole document from which emotion can be extracted (extraction models are discussed in Section 5.2).

Therefore, for each part of a document, two possible representations exist: An emotion representation (i.e., the inferred emotions) and a topical representation (i.e., the topics). The emotion representation of a part of an object combined with the corresponding topic representation defines a more complete representation of the object.

In Section 4.2, we discussed that searchers have emotion need and that it is central in the human being need system, and that searchers might not necessarily be aware of their emotion need. It was argued that underlying every information need is an emotion need, whereas the inverse is not necessarily true. This emotion need can manifest as a spectrum ranging from extreme information need to extreme emotion need. Section 4.3 discussed the idea that documents have emotion as well as topics which can be used to represent them. Finally, emotion relevance defines how the emotion aspect of a document is important to satisfy a searcher's emotion need.

4.4 Emotional Relevance in Practice - A Conceptual Map

Since the goal of an IR system is to satisfy searchers (whether informationally or emotionally), and given the preceding arguments, an IR system must then know about searchers' emotion need as well as their information need. Emotional relevance is the concept that characterises this knowledge.

This section further motivates the need for such a characterisation through discussing practical IR scenarios that can benefit from it. It is a more systematic exploration of the ways emotion relevance and emotion object can be employed for understanding IR/IS scenarios. Emotion relevance from both the searcher and system perspective are considered. Initially we explain scenarios in which the emotion part of the documents is a key factor (element) in satisfying searchers' need, then we explain how, for these scenarios, the emotion part of the documents can help in improving searchers' need satisfaction. Each perspective is explained by considering the extreme ends of the emotion need spectrum, where for simplicity extreme emotion need will just be referred to as emotion need and extreme information need as information need.

The goal of this section is to provide an exhaustive catalogue of emotion needs.

4.4.1 Searcher Perspective: The Role of Emotion Relevance in Information Need

How could the emotion aspect of a document be relevant/interesting to searchers' information need? Information need is categorised with respect to the emotion side of the documents, into two scenarios: explicit and implicit. Explicit information need scenarios are those where information need is explicitly about the emotion aspect of documents. Implicit information need scenarios are instead those where information need has an implicit relationship with the emotion aspect of documents.

Explicit Scenarios

In explicit scenarios, searchers' motivation for interacting with an IR system is to satisfy their information need, and their searchers' lack of knowledge explicitly corresponds to the emotion aspect of information object/contents. In these scenarios, the emotion inclination of the information need is usually explicitly expressed. These scenarios are further categorised into two groups: *direct* and *indirect* explicit information needs.

Direct vs indirect needs In the direct group, the emotion of the documents directly satisfies the searchers' information need: for example, searchers who seek opinions about a particular product and therefore the emotion aspect of the reviews for that product directly satisfy their IN. In the indirect-relation group, the emotion aspect of documents poses a condition on the actual information need and in that way help searchers to find relevant documents. An example from this group is a tabloid journalist who aims to find celebrity scandals.

In both direct and indirect cases the emotion aspects are important for searchers, and could be useful criteria for characterising the information need, and thereby finding relevant documents. Note that there could be topical constraints in addition to emotion ones, for better defining relevant/interested documents.

The prevalence of explicit scenarios can be explained with reference to the sociological nature of human beings, in particular with reference to the coping and empathy theories previously explained (see Section 3.5). According to the former, when searchers experience negative feelings (e.g. feelings of uncertainty), a usual response is to seek others in order to share their experience and in return learn from their experience. According to empathy theory (see Section 3.5.2), human beings are able to relate with one another through the sharing of feelings and beliefs, through which they can re-create the psychological situations of others for the purpose of understanding them (*being in their shoes*). As further discussed in Section 3.5, there are phenomena in the digital world that correspond to the empathy/coping activities in the *real* world, such as reading reviews left by other people on the web. Table 4.1 shows such scenarios.

Implicit Scenarios

In implicit scenarios, searchers' motivation for interacting with an IR system is not only to satisfy their information need, but to do so with a *latent emotional affinity*. The main difference from explicit scenarios is that searchers are either not aware of this emotional affinity, or they do not explicitly express it. Implicit scenarios are also categorised as corresponding to *direct* and *indirect* information need.

Direct In the direct case, the emotion of the document directly satisfies searchers' information need due to their latent emotion criteria in this need. A practical example of this group could be scenarios where searchers seeks news about a particular subject but only articles that, for example,

Table 4.1: Searcher information need scenarios where the emotion aspect of the contents is explicitly identified by the searcher. Each scenario is given with a corresponding example.

No.	Information need Scenario	Example
1	Search for possible range of opinions about something	Search for different emotions expressed about a new product
2	Search for a particular type of opinion about something	Search for negative emotions expressed about a new product
3	Search for a general (diverse/different/possible range of) emotion aspect of a topic	Search for the documents that talk about either advantages or disadvantages of the new health care system
4	Search for a specific set of emotions of a topic	Search for the documents that talk about the disadvantages of the new health care system
5	Search for a diverse emotion aspect of a topic	Search for different points of view on a subject
6	User explicitly requests to organise the retrieved results based on possible/defined/selected emotion categories	Organise the news of day into possible emotion categories
7	Particular emotional restriction (either type or intensity level) is applied on the retrieved results	The results retrieved for a children should not return horror items (set by parents)

are negatively inclined. A possible explanation for this behaviour could be to avoid cognitive dissonance (see Section 3.5.3).

Indirect The indirect-relation case is similar to the one for explicit scenarios, with the difference that the condition proposed here by the emotion aspect of documents on the actual information need is either hidden from the searcher or has not been explicitly expressed. An example in this group is a searcher looking for the highlights of a sporting event where the emotional features associated to highlights parts are different with respect to emotional features for the rest of the document. Table 4.2 shows such scenarios.

Table 4.2: Searcher information need scenarios where the emotion aspect of the contents is implicitly identified by the searcher. Each scenario is presented with its corresponding example.

No.	Information need Scenario	Example
1	Search for interesting contents	Searcher considers contents interesting that have been flagged as surprising content by other people
2	Search for a topic with a particular emotion inclination	Searcher considers documents as relevant that express disadvantages of the new health care system; however, the query does not express that
3	Search for important parts of an information object	Search for the highlights of a game
4	Search for novel content for a topic	User wants novel news on a subject

4.4.2 Searcher Perspective: The Role of Emotion Relevance in Emotion Need

The previous section discussed the role of emotion relevance to better answer searchers' information need. This section looks at the other end of the spectrum where emotion need is the dominant need. Here we ask the question *"how can the emotion aspect of a document be relevant to satisfying searchers' emotion need?"*

The difference from the previous section is that searchers' motivation in engaging with an IR system is to satisfy their emotional need rather than informational need. This does not mean that the information aspect of the documents is insignificant, but the focus is on the role of emotion aspect of the documents. Table 4.3 shows such scenarios.

Emotion need, like information need, is either explicit or implicit. Explicit scenarios are scenarios where searchers explicitly identify the emotion aspect of the documents they *have in mind*, whereas in implicit scenarios the emotion aspect is implicit (and could, for example, be gathered from their information history).

Table 4.3: Searcher emotion need scenarios where the emotion aspect of the contents is explicitly identified by the searcher. Each scenario is presented with its corresponding example.

No.	Emotion need Scenario	Example
1	Search for high/low arousal contents	Search for exciting/relaxing contents
2	Search for highly (and differently) commented contents	Search for controversially commented contents
3	Search for highly (but particular) commented contents	Search for contents people greatly loved/were surprised by/etc.
4	Search for a particular emotion content	Search for funny clips

The prevalence of explicit scenarios can again be explained by coping and empathy theories (see Section 3.5). The empathy part of the explanation is similar to that for explicit information need in Section 4.4.1, but the coping part differs in terms of strategies at hand. That is, in an explicit information need, the coping strategy is *problem-focused* whereas in explicit emotion need it is *emotion-focused* (see discussion of these strategies in Section 3.5.3).

In implicit scenarios, the emotion need has not been explicitly identified by the searchers. The interaction history often indicates a particular emotion affinity. As is the case for explicit scenarios, the purpose of interaction is mainly hedonic. Table 4.4 shows such scenarios.

Other possible motives for searchers to have an emotion need can be explained by *mood management* theory (see Section 3.5.4). The underlying assumption of these theories is that searchers are not necessary aware of their motives for exhibiting hedonic behaviour. These theories can thus be a good fit for emotion need in general, for both explicit and implicit scenarios. The discussion in Section 3.5.4 with regard to other possible theories explaining hedonic-based retrieval and seeking behaviour, such as *disposition theory*, are all relevant for explaining search behaviours/strategies for satisfying emotion needs.

Emotion needs of searchers have similar challenges to that of information need (in general), including an accurate understanding and expression of the need. These issues have been investigated for decades in IR community and, as a result, approaches to solve them have appeared, as for example, pseudo-relevance feedback, diversity, personalisation, etc.

The next section discusses, from a system perspective, how emotion relevance could be incorporated into an IR system to better satisfy needs. This perspective is investigated by considering what a system knows from a searcher that is, regardless of their need type (i.e. emotional or informational), either explicit or implicit.

Table 4.4: Searcher emotion need scenarios where the emotion aspect of the contents is implicitly identified by the searcher. Each scenario is presented with its corresponding example.

No.	Emotion need Scenario	Example
1	Search for interesting contents	Music that has been flagged as surprising content by other people
2	Search for contents with a particular emotion inclination	Movies that have a happy ending
3	Search for contents with a particular emotion intensity	Music that is energetic
4	Search for a query in context (Mood)	Restaurants that are cosy

4.4.3 System Perspective

This section presents a more general view of the usefulness of the emotions extracted from a document for an IR system. We start by explaining the role of emotion relevance in responding to searchers who explicitly identify their emotion criteria. It might be assumed that for such scenarios the task of IR and seeking systems is fairly easy, in a sense, since what they have to do is to match the emotion aspect of document with the defined criteria. However, the case is more complex.

From a system point of view, there are several challenges and open questions with respect to the appropriate way of capturing these criteria, the correct way of matching them with the document emotion representation, and the efficient ways of presenting the retrieved results given the requested criteria. For example, current IR systems do not generally cater for explicitly satisfying emotion criteria. This is addressed further in Section 4.5.

For implicit information need scenarios, IR systems would have to predict whether the emotion aspect of a document could be useful to be incorporated as a criterion in their relevance judgment.

We propose two sources of evidence which IR systems can rely on in order to decide whether incorporating emotion relevance is useful. First, the characteristics and nature of the data (or datasets) could suggest a particular emotion affinity. For example, it would be possible that documents considered to be topically relevant to a particular topic be diversified with respect to their emotion aspects, and thus cover a better range of subtopics. This is further explored in Chapter 8 where we show that emotion aspects of documents can indeed have such potentiality.

Second, it is proposed that the information captured from searchers' interaction with IR systems, either as explicit or implicit feedback, stored as a profile, is an important source of evidence upon which such systems can make their decision.

If a searcher's profile suggests a correlation between what a searcher considers interesting or relevant and the emotion aspect of the visited documents (e.g. news), then the system, which incorporates emotion relevance in its relevance judgement algorithm, can more accurately and effectively retrieve or recommend documents for the searcher. The correlation here could be with documents such as reviews associated with the visited information objects (e.g. movies watched). This latter

Table 4.5: System actions to satisfy the needs when the emotion aspect is explicitly identified.

No.	Explicit need Scenario	System Action
1	Search for possible opinions on an object	Cluster/diversify opinions about the objects based on their emotions
2	Search for a particular opinion on an object	Filter opinions about the objects that match the requested emotion dimension/representation
3	Search for a general emotion aspect of a topic	Cluster/diversify items based on their associated emotions
4	Search for a specific emotion aspect of a topic	Filter objects that match the requested emotion dimension/representation
5	Search for a diverse emotion aspect of a topic	Diversify highly topically relevant documents based on their emotion dimension/representation
6	User explicitly requests to organise the retrieved results based on possible/defined/selected emotion categories	Cluster and present topically relevant documents based on their emotion categories
7	Particular emotional restriction (either type or intensity level) is applied on the retrieved results	Filter out items that match the restricted emotion dimension/representation
8	Search for high/low arousal contents	Recommend content that has an emotional intensity above/below a certain threshold
9	Search for possible highly commented contents	Recommend contents to which there are different (opposite) emotional opinions associated
10	Search for particular highly commented contents	Recommend contents to which there are particular emotional opinions (with intensity higher than a threshold) associated
11	Search for a particular emotion content	Search contents that are associated with particular emotional categories

example is extensively addressed in the context of collaborative recommendation in Chapters 6 and 7, and it is shown that emotion relevance improves the accuracy of rating prediction in such systems.

Combining emotion relevance into such systems is very dependent on the task and context. In the remainder of the thesis, we present different scenarios in which emotion relevance is successfully included in the relevance decision process, but these are just examples of many other possibilities where emotion relevance can useful.

We have discussed how emotion relevance can be practically incorporated into information retrieval and seeking systems. This was explained by considering scenarios where searchers explicitly or implicitly identified their emotion criteria. The next section presents pragmatic challenges associated with the use of emotion relevance.

4.5 Pragmatic Issues

In this section, we discuss the pragmatic challenges associated with emotion relevance. These consist of three main issues: capturing searchers' emotion need, integrating emotion relevance into other relevances, and extracting the emotions from a document.

Capturing Emotion Need As briefly discussed in Section 4.4.3, current IR/IS system interfaces do not allow searches to explicitly identify their emotion criteria. Due to the adaptation capability of human beings, their emotion needs are either translated to correspondent information needs (e.g. instead of asking for happy clips, a searcher asks for Charlie Chaplin clips) which may not be the best representation of what they want. Alternatively they add words to their query to explicitly indicate the emotion, e.g., using the phrase *"funny clips"* or *"happy clips"*. For the latter, *"happy clips"* query, if a simple IR strategy is followed, the returned documents would be those which have the highest TF-IDF value for the terms *"happy"* and *"clips"*, assuming that none of the words are in the stop-words list. It is possible that the returned results would not cover the need of the searcher.

Table 4.6: System actions to satisfy the needs when the emotion aspect is implicitly identified.

No.	Implicit need Scenario	System Action
1	Search for interesting contents	Recommend items based on the emotion expressed by other people for that content
2	Search for a topic with a particular emotion inclination	Filter content that matches the emotion aspect of the topic presented in the searcher's profile
3	Search for high/low arousal content	Recommend content that has emotion intensity above/below a certain threshold for a given topic
4	Search for high/low arousal contents	Recommend parts of an information object that have emotion intensity above/below a certain threshold
5	Search for novel content for a topic	Recommend items that have different (new) emotion representation/dimension from the ones presented in the previous searcher's interaction history for a given topic
6	Search for interesting contents	Recommend Items based on their emotional content and long-term searcher's profile
7	Search for contents with a particular emotion inclination	Recommend Items based on their emotional content and searcher's short-term browsing history
8	Search for contents with a particular emotion intensity	Recommend content that has emotional intensity above/below a certain threshold

Integrating Emotion Relevance with Other Relevances The second challenge is the integration of emotion relevance with other relevances. The effectiveness of emotion relevance depends on the domain in which it is used. One might expect that emotion relevance plays an important role in IR/IS for entertainment, but a lesser role in IR/IS for academic purposes. In general, it is expected to be topic-dependent. Therefore, the strategy for combining emotion relevance with other relevance concepts should take these factors into account.

Extracting Emotion The last but the most important challenge we discuss in this section is the possibility of extracting emotion from documents. It has been shown that emotion can be extracted from different types of documents, including multimedia content using audio and visual features, and textual contents using linguistic techniques. An example of such approaches for multimedia is the work by Kang [Kan03] which uses low-level features such as colour, motion, and shot cut rate. An example for textual content is the work by Masum et al. [SPI09] using vocabulary and grammar.

Importance of Text-based Emotion Extraction Although emotion-based interaction is more pervasive for media content, there are several important applications involving text, such as news media, opinion-based texts, advertising and text that accompanies audio/video content. A proper integration of emotion into IR/IS systems thereby starts with an investigation of the emotion content of text. This thesis opted to focus on textual documents. In the next chapter, we analyse a system we developed based on the state-of-the-art emotion extraction systems.

4.6 Chapter Summary

In this chapter, we defined the concepts of searcher's emotion need, the emotion object, and emotional relevance. Then a conceptual map for the proposed concepts was presented, explaining the searcher and system side of different IR&IS behaviour scenarios. Finally the pragmatic challenges involving the proposed conceptual map were explained.

We explained that emotion need is a central need in a searchers' need system, i.e. that there is an emotion need underlying every information need. An emotion need could be the need to diminish the negative feelings caused by the uncertainty about lacking information (i.e., extreme information need) which can be solved by finding documents containing this information. At the other extreme, an emotion need could be the need to diminish an uncomfortable emotional or arousal state such as stress (i.e., extreme emotion need), which can be solved by finding documents containing relaxing (or funny) contents.

Part III

Practical Contribution

In Part II, we described a conceptual map for IR&S process by introducing the emotion need, emotion object and emotion relevance concepts, and explaining their practical and pragmatic challenges. This part describes the practical work of this thesis: we investigated emotion extraction, the key challenge for the investigation of the role of emotion in practice, where we re-implemented a state-of-the-art textual emotion extraction system, and systematically analysed its accuracy (Chapter 5).

The effectiveness of the proposed concepts in improving the retrieval performance is investigated through three use case studies: (1) memory-based collaborative recommendation (Chapter 6) , (2) model-based collaborative recommendation (Chapter 7), and (3) search result diversification scenarios (Chapter 8).

It is important to note that this part investigates the practicality of only a small part of the proposed conceptual map presented in Section 4.4. This is due to two main reasons: (1) investigating all possible scenarios explained in the conceptual map is a long-term task; (2) some scenarios were more difficult to investigate than others due to limitations associated with them, such as lack of dataset, evaluation methodology, metrics and procedure. In particular, the use case studies investigate the practicality of two scenarios:

1. *searchers have needs (emotional or informational) which are not explicitly identified and the emotion objects can be indirectly useful* (Chapter 6 and 7)

2. *searchers have information needs that are explicitly expressed via a query but the emotion object can be used implicitly and indirectly to help users to better satisfy their need.* (Chapter 8)

Finally, similar analogy to topical relevance, we follow a bag or words approach for realisation of emotion relevance. Bag of words is not only used for text, but also it is used in image retrieval where images are represented as a bag of visual words [KL10]. So we employ this approach.

Chapter 5

Evaluation of Emotion Extraction Techniques

5.1 Introduction

Most of the potential usage of emotion in IR are only possible if emotions can be extracted from text. However, systems, and methods to evaluate them, are not easily available. In this thesis, we implemented a text-based emotion extraction system and evaluated its effectiveness.

Our emotion extraction system is based on Masum et al. [Sha08] emotion extraction system, a state-of-the-art emotion extraction technique. Their emotion extraction process involves a number of steps such as triplet extraction, sentiment analysis, and emotion analysis. However, the triplet extractor component which is critical for extracting emotion is unfortunately closed-source and restricts the use of the software. Therefore, we have re-implemented Masum et al. work, with the support of the authors.

Masum et al. system created a pre-compiled lists which provide positive and negative valence for terms from a *base list*. However, recent lexicons such as SentiWordNet (SWN) 1.0 [ES06] and 3.0 [BES10] might be more adapted for the task. In this chapter, we evaluate if they can improve the performance.

One of the difficulties leveraging emotion in IR is the complexity of emotion extraction techniques, and the lack of understanding of their effectiveness. With respect to the latter, there is neither a standard resource nor standard methodologies for evaluating the usefulness and effectiveness of emotion extractor systems. The evaluation of such systems is performed using either different resources or different metrics and comparisons are impossible to make. More importantly, many such approaches are evaluated in unrealistically ideal situations, for example, using only well-formed sentences, which are extremely infrequent in real-life language. In this chapter, we systematically evaluate our emotion extraction system with other available systems, i.e., Synesketch [Krč08] and EmpathyBuddy [LSL03].

The rest of the chapter is organised as follows. Section 5.2 provides a background by discussing the state-of-the-art approaches in emotion extraction. Section 5.3 provides a background for Masum et al. system. Our implementation of this system, called *OCC1*, along with the experimental procedure is explained in Section 5.4 and 5.5 respectively. The results and discussion are presented in Section 5.6, followed by conclusions in Section 5.7.

5.2 Literature Survey on Emotion Extraction

In general, emotion extraction techniques can be categorised into four groups: keyword spotting, common-sense knowledge, model-based, and linguistic-based. In this section, these categories are described. Finally, evaluation approaches for emotion extraction systems are presented.

5.2.1 Keyword Spotting Approach

The keyword spotting approach is the simplest way of extracting emotion from text. The assumption underlying this approach is that the text in which a term occurs belongs to a certain emotion group [Ell92]. Emotion extraction is a difficult problem due to the subjective nature of the relationship between words and emotions. However, recent works in psychology and linguistics show that some words directly refer to an emotional state (i.e., joy, fear) called *direct emotion words* and others are indirectly related to an emotional state called *indirect emotion words*. Direct emotion words, such as *"killer"* cause an emotion state in the reader. Indirect emotion words, such as *"cry"* denote a consequence of an emotional state such as cry [COF87]. However, research shows that only 4% of the words used in text are direct emotion words [PMN03]. In order to assign a probability value to indirect emotion words and expand their lexicon, or consider other information such as frequency of acquiring a term in the document and collection (i.e. tf and idf) [SVS06].

All these approaches suffer from two fundamental problems. Firstly, negational is misinterpreted, e.g *"I am happy"* and *"I am not happy"* will be annotated *happy*. Secondly, the decision made by such approaches rely on surface features of the text (spotting a direct or indirect keyword) and not on a deeper understanding of the sentence. Sentences like *"I had a car accident"*, and *"I found my house by accident"* would both be placed in the same group [Sha08], which is not desirable.

5.2.2 Common-sense Knowledge Approach

A more sophisticated emotion extractor system is the work presented by Liu et al. [LSL03][1]. This approach uses a textual affective sensing engine which utilises common-sense knowledge to classify texts into six basic Ekman emotions (see Section 3.3.2). Common-sense knowledge is represented by a graph where real-life concepts are the nodes and their relationships are the edges of the graph. A problem here is that the linguistic aspect of the sentence is not taken into account: sentences like *"it is impossible to cook a bad meal following this recipe"* and *"I will cook a bad meal following this recipe"*, will be categorised in the same group. It has also been argued that using the six categories to classify text is not optimal, as they are based on feeling theories and do not consider the cognitive aspect of emotion (such as beliefs, decisions and intentions) [Sha08] (see Section 3.2).

5.2.3 Model-based Approach

The third approach is that of model-based emotion extraction systems, based on supervised or unsupervised machine learning techniques. They build a model from an annotated set of documents and use it to categorise new documents. Examples include [WWC05], [Seb02] and [SVS07]. The effectiveness of these approaches rely on a large amount of training data, and although they perform

[1] Liu et al. [LSL03] provide an open source version of their approach called EmpathyBuddy. EmpathyBuddy considered to be the best performing open source emotion extraction system [Sha08].

well at the document or paragraph level, they do not perform well at sentence level [Sha08].

5.2.4 Linguistic-based approach

The last group are the approaches that use language processing techniques to extract emotion from text. Synesketch[2] uses the Affective WordNet lexicon [SV04] to calculate the emotion weights of words in text and categorises emotion from text into the six basic Ekman emotions. Masum et al. use contextual valence values of triplets extracted from the sentences (i.e. subject, verb, and object) and extract emotion from text using OCC model rules [SPI09], They consider cognitive science, knowledge representation, natural language processing, and contextual sentiment analysis, and their emotion extraction system is considered to be the state-of-the-art.

5.2.5 Evaluating the Approaches

There have been several proposals to evaluate emotion extraction systems. In some cases arbitrary emotion categories [AS05], or assignment of a set of emotion into one category [AS05] are used, hence making some comparisons almost impossible. For example, Alm et al. [AS05] introduced a manually annotated test collection of 22 fairy tales for the emotion extraction purpose. Sentences that express a dominant emotion were annotated using one of the eight emotion categories, namely angry, disgust, fearful, happy, sad, positively surprised and negatively surprised. One evaluation procedure was to combine all the emotional sentences into one category and all the neutral sentences to another category, to evaluate the accuracy of classifiers [AS05]. They also employed another procedure which divides the emotion category into positive and negative emotions, and evaluated the accuracy of approaches. On the other hand, Kim et al. [KVC10] used the same data set and only considered four emotion categories out of the original eight categories namely anger, fear, joy and sadness. They also followed the same procedure on the news headline data sets presented for SemEval 2007 *Affective Text* task [SM07]; news headline data sets are originally annotated using six Ekman emotion categories, therefore considering only four of these six

[2]Synesketch is an open-source Java API for textual emotion recognition. The API is bundled with a number of imaging tools that allows users to automatically create visualisations based on textual emotions. More information can be found at www.synesketch.krcadinac.com.

categories could deem Kim et al. results arbitrary and inconclusive.

Given the need to exploit emotion in IR for various applications, it is important to understand and quantify the quality of emotion extraction techniques. In this chapter, we provide a systematic approach for evaluating emotion extraction systems, that attempts to remedy the lack of standardisation in the current approaches.

5.3 Masum et al. Approach

We based our emotion extraction system on Masum et al. [Sha08], and hence further describe their approach in this section. Figure 5.1 shows the architecture of Masum et al. [Sha08] system and the components involved, namely the prior valence provider, triplet extractor, sentiment analyser, and emotion extractor. Each is detailed below.

5.3.1 The Prior Valence Provider

This component is responsible for creating and expanding a set of base lists, each of which maps set of words (categorised morphologically, e.g., verbs, adverbs, adjectives, and nouns) to their prior valence (i.e., positive or negative). The initial base lists of verbs, adjectives and adverbs are created with the help of WordNet [Mil95]. A base list for nouns is created with the help of ConceptNet and an initial base list for named entities is also formed using Opinmind[3]. If a prior valence for a term is not available in a base list, the prior valence provider automatically assigns a valence for that word by first obtaining the synonyms of that word using a thesaurus[4], then screening the synonyms with respect to the corresponding base lists for which numerical values are already assigned, and finally averaging the obtained valence as the valence value of the word. The new word and its valence are then inserted into the base list.

[3] www.Opinmind.com
[4] www.thesaurus.com

5.3.2 The Triplet Extractor

A triplet refers to the three-component structure of a sentence: subject, verb and object. Each component may have several attributes called modifier. Examples of modifiers are adjectives, adverbs, and noun phrases. To extract triplets, first a set of triplets for a given sentence is obtained from the Machinese Syntax[5] (i.e. syntactic parser). Since a sentence can consist of several sub-sentences, the parser can extract more than one triplet per sentence. In this case, there would be a dependency between the extracted triples, based on the dependency between the sub-sentences of a given sentence.

5.3.3 The Sentiment Analyser

All the triplets obtained from the input sentences are processed to assign a valence value to the sentence. The valence of the attributes is calculated by averaging the prior valence value of the terms in the attribute set. Then the valence of each part (i.e. subject, verb, or object part) is calculated by applying a set of rules considering the valence value of the actual part (i.e. subject, verb, or object) and its attribute part. Then the valence of a triplet is calculated by first combining the valence of verb and object part (named as verb-object valence) and then combining the valence of subject part with the verb-object part. The combination is based on a set of rules [SPI08]. Finally, the overall valence of the sentence is calculated [SPI08] based on the type of dependencies among triplets.

5.3.4 The Emotion Extractor

The cognitive structure of the OCC model can be characterised by specific rules and their interplay with several variables. There are two kinds of variables involved: emotion inducing variables and emotion intensity variables. Multiple emotions can be inferred from a given situation depending on whether states expressed by certain cognitive variables hold or not hold.

[5] www.connexor.com/connexor/

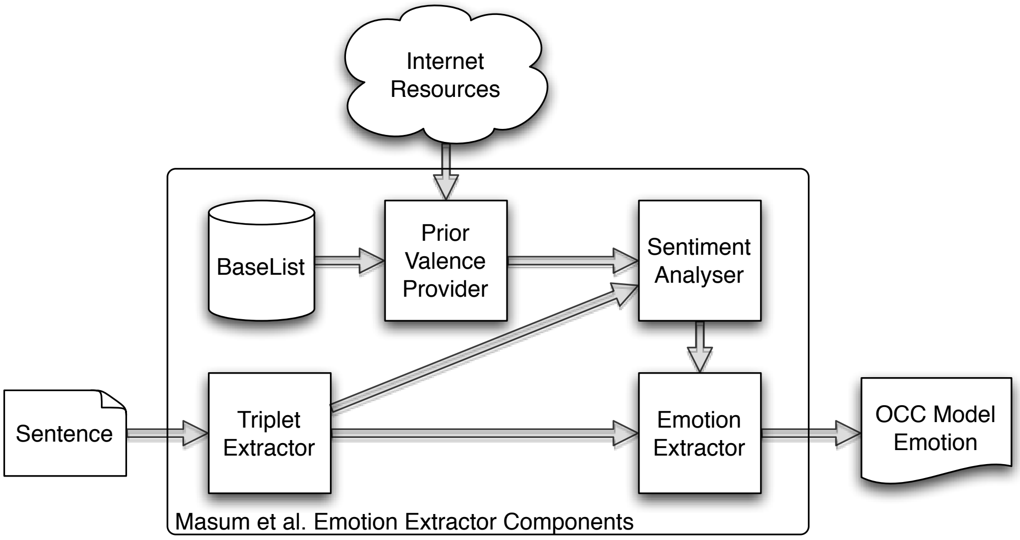


Figure 5.1: Masum et al. Emotion Extractor Architecture

5.4 Implementation of Masum et al. Approach

In this section, we explain our implementation of the Masum et al. [Sha08] approach which we refer to as *OCC1*. Only the differences with the original approach are discussed.

5.4.1 The Prior Valence Provider

The prior valences of terms in the base lists is a critical component of the emotion extractor. In order to study the effectiveness of more recent lexicons in the accuracy of the system, Three sets of base lists from diverse lexicons namely General Inquirer (GI) [SDSO68], SWN 1.0 [ES06] and 3.0 [BES10] were created.

GI provides *positive*, *negative*, *strong*, *weak*, and *othertags* features for 11790 terms. In order to create the base lists from GI, these features were considered. The *othertags* feature provides the morphological information of a term which is required to decide which base list (i.e. adjective, adverb, verb or noun list) the term belongs to. The *strong*, and *weak* features indicates the strength of the valence value. Each term gets a valence value of one if the term is judged as weak, two if there is no judgement available, and three if the term is judged as strong. The *positive* and *negative* features indicate the sign of the valence value. If a term was judged neither as positive nor as negative, then the value for it will be 0 regardless of the strong and weak features.

SWN 1.0 provides positive and negative scores between zero and one for 115420 terms with the *POS* feature providing the morphological information. The SWN 3.0 provides the same information for 117660 terms. In order to create base lists from SWN 1.0 and 3.0, the difference between the negative and positive synset is calculated. For comparison, we also consider the base lists provided by Masum et al. and we refer to them as *Base List 1*. The effectiveness of the prior valence provider also depends on the expansion of the base lists. For the expansion of the base lists we follow the procedure described in Section 5.3.

5.4.2 The Triplet Extractor

The triplet extractor developed in Masum et al. work depends on a closed-source syntactic parser application. In order to overcome this problem two triplet extractors were developed, namely Stanford-based triplet extractor and ConceptNet-based triplet extractor. The Stanford-based triplet extractor is based on stanford parser which generates a highly accurate (86.32%) parse tree[6] for a given sentence [KM03, cDMMM06] and is an open-source resource.

ConceptNet-Based Triplet Extractor is based on ConceptNet which is also an open-source package providing set of functions including generating its parse tree. A drawback of this package is that there is no information available on its accuracy.

The Stanford-based Triplet Extractor

The Stanford-based Triplet Extractor utilises the output of the Stanford parser. Given a sentence, Stanford parser returns a parse tree. Figure 5.2 shows the parse tree for the sample sentence *"relentless optimisation of information retrieval effectiveness has driven web search engines to new quality levels where most people are satisfied most of the time."*

To extract triplets from the parse tree, the deepest sentence (DS) in the tree is first found. DS is the sentence which does not have any other sentences under itself in the parse tree (shown in Figure

[6]Parse tree is a tree that represents the syntactic structure of a sentence based on a formal grammar.

5.2). Each DS usually consisting of two main parts, the noun phrase (NP), and verb phrase (VP). The deepest noun in the NP is considered as the subject of the DS and other members in the NP such as adjectives and nouns are considered as the attributes of the subject.

The VP usually consist of two parts, namely verb and object. The deepest verb in the VP is considered as the verb of DS, and adverbs in DS are considered as the attributes of the verb. The deepest noun in the VP is considered as the object of the DS. As before, other members in the NP such as adjectives and nouns are considered as the attributes of the object. The information extracted from each DS constitutes a triplet. After that the DS is removed from the sentence and the procedure is repeated until there is no DS left in the sentence. Finally, the dependencies among triplets are taken into account.

```
(ROOT
  (S
    (NP
      (NP (JJ relentless) (NN optimization))
      (PP (IN of)
        (NP (NN information) (NN retrieval) (NN effectiveness))))
    (VP (VBZ has)
      (VP (VBN driven)
        (NP (NN web) (NN search) (NNS engines))
        (PP (TO to)
          (NP
            (NP (JJ new) (NN quality) (NNS levels))
            (SBAR
              (WHADVP (WRB where))
              (S
                (NP (JJS most) (NNS people))
                (VP (VBP are)
                  (VP (VBN satisfied)
                    (NP
                      (NP (JJS most))
                      (PP (IN of)
                        (NP (DT the) (NN time)))))))))))
    (. .)))
```

Deepest Verb

Deepest Sentence

Figure 5.2: A sample output of the Stanford parser for the sentence: *"relentless optimisation of information retrieval effectiveness has driven web search engines to new quality levels where most people are satisfied most of the time."*

ConceptNet-based Triplet Extractor

Given a sentence, ConceptNet [LS04] returns its parse tree. ConceptNet uses a graph of concepts which is built on a large semantic networks of common sense knowledge [LS04]. Given Concept-

Net output (i.e., the parse tree), we create the triplets similar to the procedure explained in previous section.

5.4.3 The Sentiment Analyser

The only difference in our implementation with respect to the original approach is that we consider two methods for combining the valence of the attributes instead of one. In the first method, the valence of the attributes is calculated by averaging the absolute value of the valences of the terms in the set. If there is one negative valence value in the set, the sign of the valence will be negative. The reason behind this decision is that we believe negative words have more influence than positive words, and that the use of one negative word can be a strong indication of negative valence for those attributes. Another approach is simply to average the valences in the list. This approach is used by Masum et al.

The same consideration is applied when combining the triplets. Two approaches are used. The first is the method reported in Masum et al. (i.e if the number of negative triplet is bigger than the case of positive triplets, the sentence is considered to be negative). The second considers the whole sentence negative when finding one negative triplet within it.

5.4.4 The OCC Emotion Extractor

This part exactly follows the procedure explained in Masum et al. [SPI09] work (see Section 5.3.4).

In addition to OCC1 system, two publicly available commercial systems, Synesketch [Krč08] and EmpathyBuddy [LSL03] were used. Next section explains the methodology used for evaluating their performance.

5.5 Experimentation Methodology

In this section, we first present the experimental condition in Section 5.5.1, Followed by the evaluation metrics are presented in Section 5.5.2. The data sets are described in 5.5.3, then the threshold problem for the News Headline data set is explained in Section 5.5.4. Finally, the evaluation procedure is described in Section 5.5.5.

5.5.1 Experimental Conditions

A number of conditions were defined to test the components and factors involved in OCC1 system. The assumption is that recent lexicons should improve the effectiveness of both the sentiment analyser and emotion extractor systems with respect to the base lists Masum el al. provided. Another assumption is that using an *external resource* (i.e. thesaurus.com) to expand the base lists should improve the effectiveness of OCC1.

The effectiveness of the sentiment analysis depends on the *triplets* and how to *combine* their valence values. It also depends on the *valence of attributes* (i.e. *adjective* and *adverbs*) of the subject, object, verb. Contextual rules such as *negation* and *dependencies* among triples are also considered to be important factors. Our assumption is that the best performing approach for both sentiment analyser and emotion extractor is the one using all these features. All these variations (highlighted with italic) amount to 8 conditions and the combinations of them generates 128 (2^8) settings. Each setting is then repeated four times with our four base lists. Note that the approaches are rule-based, requiring no training.

5.5.2 Metrics

Effectiveness of systems were compared using precision, recall, and F-measure. These measures are commonly used in sentiment analysis and emotion extraction literature. For sentiment analysis, the results are calculated for positive sentences, negative sentences and overall. It was important to find an approach that has both the highest overall effectiveness and the best balanced effectiveness in both positive and negative sentences. For emotion extraction, the results are calculated for each individual emotion as well as overall.

5.5.3 Data Sets

We now describe the different data sets available for evaluating the text-based sentiment and emotion extraction systems which are Sentence Polarity v1.0 and News Headlines data set respectively.

Sentence Polarity Data Set v1.0:

This data set contains 5331 positive and 5331 negative processed sentences and snippets, which were introduced by Pang and Lee in 2005 [PL05]. It was used to investigate the experimental condition described in Section 5.5.1, and is ideal for evaluating the effectiveness of a sentence level sentiment extractor system.

Other sentiment analysis data sets assign a positive or negative value to a document, such as the Blog06 Trec collection [MO06]. These data sets, however, complicate the evaluation since they provided the valence value to the documents rather than to the sentences, thus we ignored these data sets.

News Headlines Data Set

For our experiments we used the collection *SemEval-2007 Task 14: Affective Text* [SM07] consisting of a training set which contains 250 headlines and a test set which contains 1000 headlines. Each news headline is annotated with six Ekman emotions. For each emotion, an interval between 0 to 100, where 0 indicates that the emotion is not present in the given headline, and 100 indicates maximum emotional load, is used.

These data sets are suitable since (1) news headlines are intentionally written with an emotionally rich content to provoke readers' attention [SM07]; (2) the emotion extractor systems that we are examining are sentence-based level; and in addition, (3) the outcome of two benchmarks (i.e., Synesketch and EmpathyBuddy) are Ekman emotion categories. The outcome of Masum et al. is not compatible with the Ekman emotion model and there was no guidance in the literature available on how the OCC model's emotions map to Ekman emotions. We therefore did a semantic mapping presented in Table 5.1. In addition, there are no publicly available data sets with OCC model emotion categories.

OCC Model Emotion	Ekman Emotion
Surprise, Shock	Surprise
Hate, Anger, Resentment	Anger
Fear, Fears-confirmed	Fear
Sorry-for, Distress, Remorse, Shame	Sadness
Joy, Happy-for, Gloating, Relief, Pride, Admiration, Love, Gratification, Satisfaction, Gratitude, Hope	Joy
Reproach, Disappointment	Disgust

Table 5.1: Mapping of OCC model emotion to Ekman emotion

5.5.4 News Headlines Data Set Threshold Problem

OCC1 is rule-based, and hence its output is a binary value, corresponding to the dominant emotions in the sentence. Since we would like to compare the performance of this system with other approaches, and also study the effect of each feature in OCC1, coarse-grained metrics such as precision, recall and F-measure are suitable as opposed to fine-grained metrics such as correlation measures.

In Strapparava and Mihalcea [SM07], a threshold of 50 is introduced in order to transform the emotion value provided by the judges for each emotion in each news headline sentence to a binary scale. All the emotions with the value in the range of (50, 100] will be considered as 1 and those in the range [0, 50] as 0. However, in our opinion, this threshold is too high, having a negative impact on the performance of OCC1, and gives us a wrong estimation of the effectiveness of this

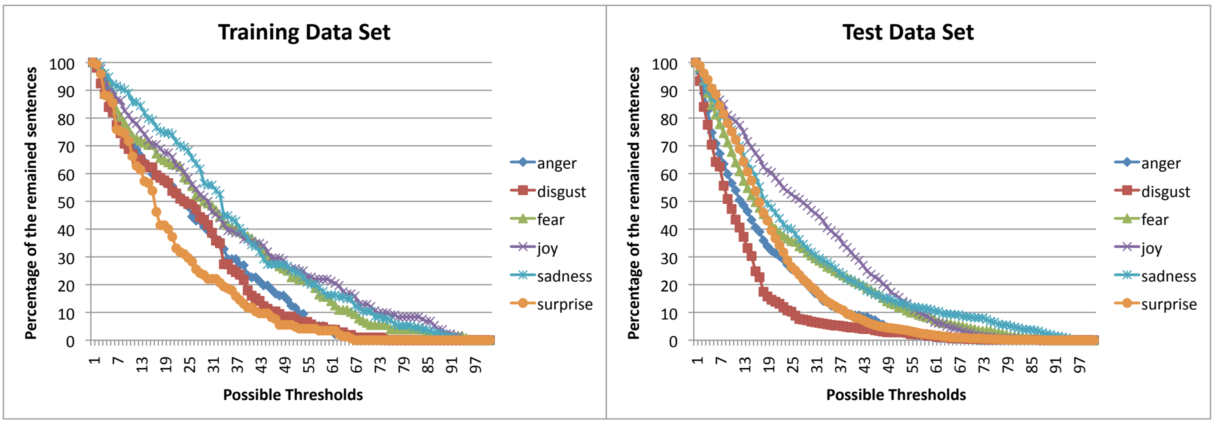


Figure 5.3: The percentage of remained sentences with emotion at each threshold point for training and test sets

system. Figure 5.3 shows that over all emotions, on average, for the training set is more than 80%, and for the test set, more than 90% of the emotion values are below threshold 50. On the other hand, considering the threshold 0 might not be a good solution neither as two sentences that have been assigned an emotion value of 1 and 99 would be treated equally.

In order to overcome the threshold setting problem, an exhaustive experiment considering all possible thresholds from 0 to 100 was performed. The system that outperforms other approaches for a higher number of thresholds will be considered as the best performing system. Finally, all systems were tuned using the training set and tested on the test set.

5.5.5 Evaluation Procedure For the News Headlines Data Set

Tuning EmpathyBuddy and Synesketch

The output of EmpathyBuddy and Sysesketch is a value from 0 to 1 for each of the Ekman emotions. Since we introduce a threshold on the test set, we also have to optimise the effectiveness of these systems with respect to the threshold. Not defining a cut-off value for each emotion for these systems introduce a bias in the comparison of the systems as the precision value of these systems will be lower than their optimal value. Therefore, it is important to calculate the cut-off value. This should be the value where the system has the highest F-measure. We, thus, selected, for each emotion, a different cut-off value corresponding to the highest F-measure value on the training set.

Tuning OCC1

As it is explained in the experiment section, there are 512 possible combinations of factors that each resulting in different performance of OCC1. The training set was used to find the best possible combination of features. The decision was based on the F-measure value and the best performing combination was used on the test set.

5.6 Results

In this section, we first present the best performing emotion extraction system across all emotions in Section 5.6.1, followed by the results obtained for the sentiment analysis component on the Sentence Polarity data set in Section 5.6.2. Finally, the main and interaction effects of the parameter settings of the OCC1 performance is described in Section 5.6.3.

5.6.1 Best Performing Approach Across All Emotions

An evaluative comparison between the Synesketch, EmpathyBuddy and OCC1 emotion extractors across all emotions is presented in Figure 5.4, 5.6 and 5.5. In each figure, the x-axis corresponds to the thresholds applied on the test collections (i.e. 0 to 100) and the y-axis to the value of the metric (i.e. precision, recall, and F-measure). The points in the graph represent the performance of the systems on the test set. Each system used optimised parameters tuned on the training set. For EmpathyBuddy and Sysesketch, the optimum is the best cut-off value for each emotion and for OCC1, the optimum is the best combination of the factors explained in section 5.5.1.

For OCC1, the performance of the system is more sensitive to base list (i.e. GI, BaseList1, SWN 1.0 and SWN 3.0) and the base list expansion (i.e. using Thesaurus or not) factors than to the other factors explored. There are two reasons for this: (1) these factors together provide the base knowledge for OCC1; (2) these factors are external to the system and there is no control over them. In order to provide a comprehensive study, it is important to understand their effect on the effectiveness of OCC1. Thus, for each threshold applied on the test collection, eight outcomes of

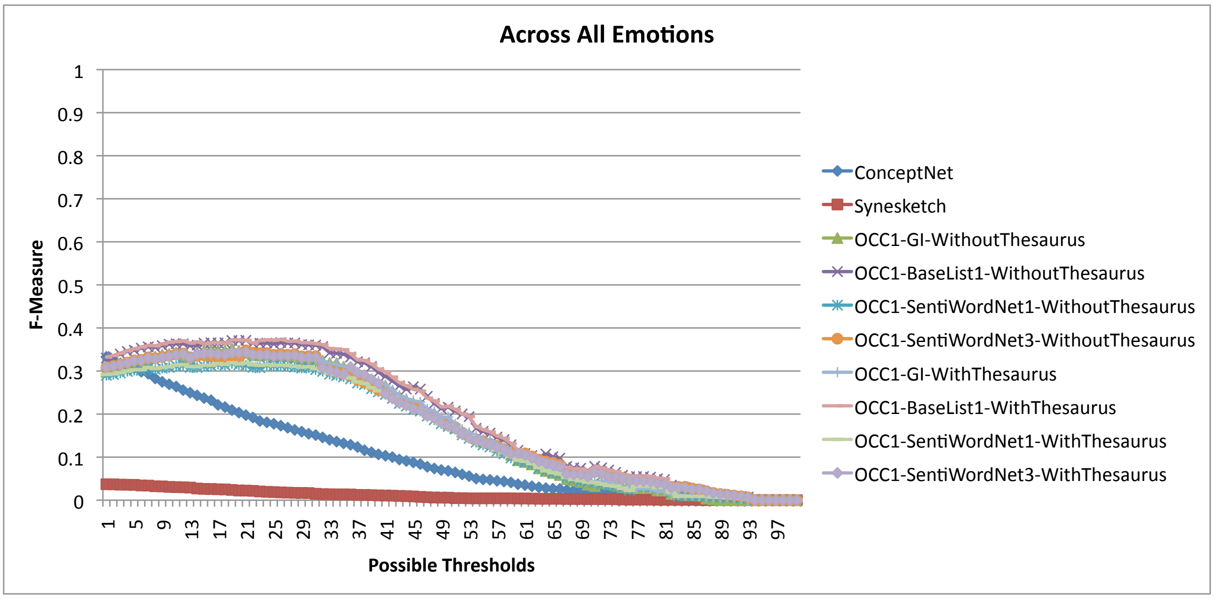


Figure 5.4: F-measure values across all emotions for all the thresholds on the news headline test set

OCC1 with respect to the possible combinations of these factors are presented in Figure 5.4, 5.5, and 5.6.

As shown in Figure 5.4 and 5.5, OCC1 is more accurate in extracting emotion from text for the F-measure and precision, across all emotions and thresholds, whereas Figure 5.6 shows that EmpathyBuddy is better in terms of recall. The reason for the lower recall of OCC1 compared to EmpathyBuddy is (1) that OCC1 provides the dominant emotions in the sentence rather than probability values for the six Ekman emotions; and (2) OCC1's output is mapped from the OCC model to the Ekman emotion model, which is error prone.

Sysesketch has the lowest effectiveness in comparison to other systems in terms of precision, recall, and F-measure. This is probably due to the lack of accuracy of the Affective WordNet base list used in this system and/or the naive linguistic interpretation of the sentences.

For OCC1, across all emotions and thresholds, using the *Base List 1* performed the best. The SWN 3.0 list performs better than SWN 1.0 list. This result is consistent with other results on these two lexicons [ES09]. One reason is that SWN 1.0 is based on WordNet 2.0 whereas version 3.0 is based on a newer version of WordNet (version 3.0) which covers more words. Another reason is

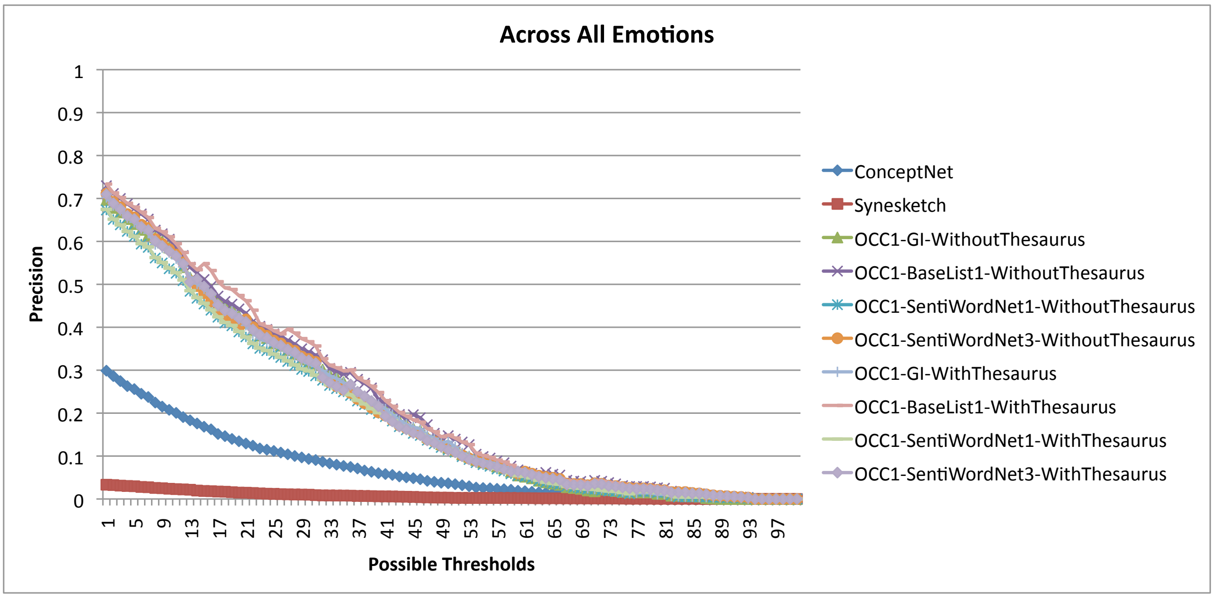


Figure 5.5: Precision values across all emotions for all the thresholds on the news headline test set

that SWN 1.0 is constructed based on a semi-supervised learning algorithm while version 3.0 goes one step further and uses an iterative random-walk process.

The effectiveness of OCC1 in terms of precision and F-measure is very close when using either GI or SWN 3.0. For high thresholds, OCC1 using SWN 3.0 has a higher recall than OCC1 using GI, possibly since GI covers a smaller number of terms than SWN 3.0. In other cases they have a similar effect. In terms of precision and F-measure, we observe a consistent behaviour across all variations of OCC1. The differences between each base list were more apparent in the recall metric when the thresholds went above 55.

As shown in Figure 5.4 and 5.5 the use of a base list expansion[7] did not enhance the recall value of OCC1 using SWN 1.0 and 3.0, but increased the recall value of OCC1 using GI, possibly since the information in *thesaurus.com* is mostly presented in the SWN data sets, since there is a large number of terms contained. Therefore, using synonyms do not bring any new information whereas in GI base lists this is not the case. In general, expanding the base lists using a thesaurus slightly improves the F-measure value. However, this improvement is not consistent over all thresholds and the improvement depends upon base list.

[7]using thesaurus.com

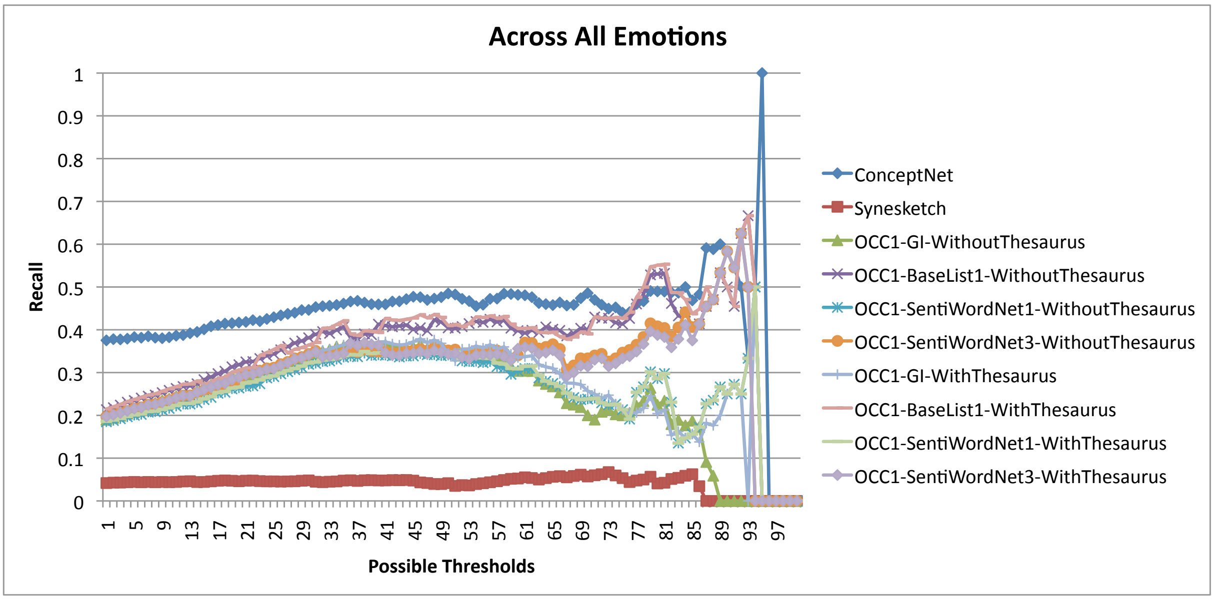


Figure 5.6: Recall values across all emotions for all the thresholds on the news headline test set

5.6.2 Sentiment Analyser Performance on the Sentence Polarity Data Set

This section examines the accuracy of the sentiment analyser system of OCC1, presenting results on the sentence polarity data set for settings explained in the experimental condition section (Section 5.5.1). Results are reported in Table 5.2, where columns correspond to the base lists and rows to the situations where the experiments are either conducted with or without base list expansion.

OCC1 using Base List 1 performed better than using other lexicons. OCC1 using SWN 3.0 performed better than OCC1 using SWN 1.0 list. As shown in Table 5.2, the base list expansion did not enhance the results of settings using SWN 1.0 and 3.0, whereas it increased the recall value of the settings using GI base lists.

	SentiWordNet 1.0	SentiWordNet 3.0	General Inquirer	Base List 1
Without	0.548, 0.532, 0.54	0.554, 0.539, 0.546	0.58, 0.45, 0.508	**0.592, 0.584, 0.588**
With	0.547, 0.532, 0.54	0.552, 0.542, 0.547	0.555, 0.543, 0.549	**0.597, 0.592, 0.595**

Table 5.2: Average of precision, recall, and F-measures on the positive and negative samples in the sentence polarity data set v1.0

Features of the Best Performing Setting

In the preceding, it has been shown that the highest value of F-measure among all systems across all emotions corresponds to OCC1, and among different OCC1 settings the one which uses the *base list 1* with expansion performs better than others. In Figure 5.4, we can see that for this setting, OCC1 has an optimal F-measure value for the threshold 26 on the test set. Further investigation on this setting at this threshold shows that the combination that utilises adjective, adverb, negation, and contextuality outperformed combinations in which one or more of these components were absent. This combination of features was consistent with the combination of features performed best in sentiment analyser system, implying that the effectiveness of the emotion extractor directly depends on the effectiveness of the sentiment analyser system.

5.6.3 Feature Analysis of OCC1

This section reports extensive experiments on the effect of parameters on OCC1 performance (see Table 5.3). This is important since OCC1 is not a machine learning approach and hence we need to exhaustively study their effect. The effectiveness of the system is reported with the F-measure, for each possible setting of the parameters. F-measure was chosen as it gives a synthetic value that combines precision and recall. The main and interaction effects of these features are reported.

Variable Name	**Possible Choices**
Using Thesaurus	With - Without
Base List	BaseList1 - GI - SentiWordNet1.0 - SentiWordNet3.0
Triplet Extractor Method	TripletExtractor - TripletExtractorBasedOnConceptNet
Contextual Valence Negation	Negative Bigger than Positive - One Or More Negative
Attributes Combination Method	Average - Not Average
Using Adjective	With - Without
Using Adverb	With - Without
Using Negation	With - Without

Table 5.3: Possible choices for the features considered in the experiment

Main Effect Analysis

In following, each table shows the mean and standard deviation of the F-measure value when the feature is either used or not. For comparison, we report the effect of each feature for four thresholds: 1 (Highest Recall), 26 (Optimal), 50 (Middle) and 80 (Highest Precision); with each row corresponding to one of the four. Two-Way Anova was applied on each feature to examine its statistical significance (the tests with $P < 0.05$ indicated by *, $P < 0.01$ indicated by **, and $P < 0.001$ indicated by ***).

Table 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 5.10, 5.11, and 5.12 show the mean and standard deviation for the F-measure considering the effect of *using thesaurus*, *base list*, *triplet extractor method*, *using attributes*, *contextual valence negation*, *contextual valence negation*, *using adjective*, *using adverb* and *using negation* respectively.

Using Thesaurus The results indicate that using the thesaurus improves the performance of the systems. However, this improvement is not statistically significant, since base lists show different behaviour when expanded. For instance, we have seen before that SWN 1.0 and 3.0 did not benefit from using Thesaurus whereas Base List 1 did.

Using Thesaurus	With	Without
Threshold	M (SD)	M (SD)
1	0.301 (0.02)	0.298 (0.019)
26	0.33 (0.026)	0.329 (0.024)
50	0.192 (0.016)	0.191 (0.016)
80	0.031 (0.012)	0.030 (0.014)
Polarity Data Set	0.504 (0.047) *	0.497 (0.057) *

Table 5.4: Threshold setting, mean and standard deviation for the F-measure considering the effect of *using thesaurus*

Base Lists The results indicate the different effects each base list has on system effectiveness are statistically significant. The results show that *Base List 1* performs the best across all thresholds whereas the SWN 1.0 performs the worst.

Base List	Base List 1	GI	SentiWordNet1.0	SentiWordNet3.0
Threshold	M (SD)	M (SD)	M (SD)	M (SD)
1	0.324 (0.009) ***	0.295 (0.013) ***	0.281 (0.014) ***	0.299 (0.011) ***
26	0.368 (0.004) ***	0.327 (0.006) ***	0.3 (0.009) ***	0.324 (0.008) ***
50	0.213 (0.006) ***	0.185 (0.007) ***	0.176 (0.005) ***	0.193 (0.014) ***
80	0.046 (0.004) ***	0.016 (0.007) ***	0.021 (0.005) ***	0.038 (0.005) ***
Polarity Set	0.556 (0.013) ***	0.432 (0.048) ***	0.503 (0.019) ***	0.513 (0.017) ***

Table 5.5: Threshold setting, mean and standard deviation for the F-measure considering the effect of *base lists*

Triplet Extractor Type The results indicate that Stanford-based triplet extractor performs better for low thresholds whereas ConceptNet-based triplet extractor performs better for the high thresholds (i.e. dominant emotions), with statistical significant effectiveness differences. The reason may be due to the fact that the triplet extractor returns more terms as attributes of triplet components (i.e. subject, verb, and object), hence, improving performance for emotion extraction, hence increasing the recall at the cost of precision.

Triplet Extractor Method	TripletExtractor	TripletExtractorBasedOnConceptNet
Threshold	M (SD)	M (SD)
1	0.311 (0.014) ***	0.288 (0.018) ***
26	0.334 (0.021) ***	0.326 (0.028) ***
50	0.185 (0.015) ***	0.199 (0.014) ***
80	0.026 (0.013) ***	0.035 (0.012) ***
Polarity Set	0.509 (0.047) ***	0.493 (0.047) ***

Table 5.6: Threshold setting, mean and standard deviation for the F-measure considering the effect of *triplet extractor method*

Using Attributes Type The results indicate that there is not much performance difference in considering this feature. This is possibly related to the news headline data set since each news item in this data set is concise and so it usually doesn't have several adjectives or adverbs. This data set is therefore not suitable for this feature. The results on the sentence polarity data set show that averaging of the attributes provides better effectiveness than the situation in which negative sign has priority.

Using Attributes	Average	Not Average
Threshold	M (SD)	M (SD)
1	0.3 (0.019)	0.299 (0.02)
26	0.33 (0.025)	0.33 (0.025)
50	0.192 (0.016)	0.192 (0.016)
80	0.03 (0.013)	0.03 (0.013)
Polarity Set	0.504 (0.054)	0.498 (0.051)

Table 5.7: Threshold setting, mean and standard deviation for the F-measure considering the effect of *using attributes*

Contextual Valence Type The results indicate that there is not much difference in performance when considering the feature. This is possibly due to the conciseness of news items as they may not consist of several sub-sentences or clauses. Therefore, this data set may not be the suitable for this feature. The results on the sentence polarity data set show that the situation where one triplet is negative leads all triplets to be considered negative performs better than the other setting.

Contextual Valence Negation Type The results indicate that there is not much difference in performance when considering the feature and in similarity to the previous feature, the same data set problems have been encountered. The results on the sentence polarity data set show that the situation where one triplet is negative leads all triplets to be considered negative performs better than the other setting.

Contextual Valence	False	True
Threshold	M (SD)	M (SD)
1	0.2 (0.02)	0.2 (0.02)
26	0.33 (0.025)	0.33 (0.025)
50	0.192 (0.016)	0.192 (0.016)
80	0.03 (0.013)	0.03 (0.013)
Polarity Set	0.497 (0.051) *	0.505 (0.054) *

Table 5.8: Threshold setting, mean and standard deviation for the F-measure considering the effect of *contextual valence type*

Contextual Valence Negation	NegBiggerPos	OneOfMoreNeg
Threshold	M (SD)	M (SD)
1	0.299 (0.02)	0.299 (0.02)
26	0.33 (0.025)	0.33 (0.025)
50	0.192 (0.016)	0.192 (0.016)
80	0.03 (0.013)	0.03 (0.013)
Polarity Set	0.499 (0.053)	0.503 (0.052)

Table 5.9: Threshold setting, mean and standard deviation for the F-measure considering the effect of *contextual valence negation type*

Using Adjective Type The results indicate that there is no unanimous behaviour across different thresholds. This indicates that this feature is sensitive to the input sentences. We believe that the performance of this feature was affected by other features such as base list, thesaurus, negation and triplet extractor. The next section examines the interaction effect of using adjective and other features to understand this effect better.

Using Adverb Type The results indicate that there is not much difference in system performance when considering the feature. This is possibly due to the fact that each base list performed differ-

Using Adjective	With	Without
Threshold	M (SD)	M (SD)
1	0.303 (0.019) ***	0.296 (0.019) ***
26	0.33 (0.024)	0.329 (0.026)
50	0.19 (0.016) ***	0.194 (0.016) ***
80	0.031 (0.01)	0.03 (0.015)
Polarity Set	0.52 (0.034) ***	0.482 (0.061) ***

Table 5.10: Threshold setting, mean and standard deviation for the F-measure considering the effect of *using adjective type*

ently with respect to the adjective, and since using negation had a strong effect on the performance of the system. It is important to see the effect of this feature with respect to other features, and next section examines the interaction effect of using adjective and other features to understand this effect better.

Using Adverb	With	Without
Threshold	M (SD)	M (SD)
1	0.3 (0.019)	0.299 (0.02)
26	0.33 (0.025)	0.33 (0.025)
50	0.192 (0.016)	0.192 (0.016)
80	0.03 (0.013)	0.03 (0.013)
Polarity Set	0.504 (0.051)	0.498 (0.054)

Table 5.11: Threshold setting, mean and standard deviation for the F-measure considering the effect of *using adverb type*

Using Negation Type The results indicate that system performance does not change much considering the feature. It is possible that this is due to not having many sentences with negation in the data set. Thus this data set may not be suitable for this feature.

Using Negation	With	Without
Threshold	M (SD)	M (SD)
1	0.3 (0.02)	0.299 (0.019)
26	0.33 (0.025)	0.329 (0.025)
50	0.192 (0.016)	0.192 (0.016)
80	0.03 (0.013)	0.03 (0.013)
Polarity Set	0.502 (0.053)	0.5 (0.052)

Table 5.12: Threshold setting, mean and standard deviation for the F-measure considering the effect of *using negation type*

In summary, the main affect analysis of the features on the F-measure performance of OCC1 indicate that the base list and triplet extractor have great impact on the performance. Other features did not show any differences of their own and need to be checked in interaction with other features.

Interaction Effect Analysis

The interaction effect between using thesaurus and different base lists for threshold 1 and 26 and 80 indicates that there is a statistically significant effect between these two features whereas this effect is not significant for the threshold 50. There is also an effect between the base list and triplet extractor type, and between the base list and using the adjective feature. Comparison of multiple features indicates that there is a statistically significant effect between following features: thesaurus, base list, triplet extractor, negation, adjective, and adverb. The experiment did not show any statistically significant effect for contextuality and negation for contextuality.

On the sentence polarity data set, Stanford-based triplet extractor outperformed ConceptNet-based triplet extractor. This is possibly due to the fact that Stanford-based triplet extractor considers more syntactical aspects of the sentence. After the triplet extractor, contextuality and adjective were the most influential features in the effectiveness of the system. Analyses of the features show that contextuality and negation for contextuality have statistically significant effects on each other, and the thesaurus has a statistically significant effect with base lists. In summary, the results obtained from the interaction effect analysis show that the OCC1 which employ all features results in the best performing system.

5.7 Chapter Summary

This chapter presented a comparative study of text-based emotion extraction techniques. Our study showed that our implementation of emotion extraction method proposed in [Sha08], called OCC1, is more accurate in terms of precision and F-measure. Also the effect of various base lists in the performance of OCC1 was analysed, showing *Base List 1* to be the best, and SentiWordNet 3.0 to be better than SentiWordNet 1.0. Finally, the influence of each of the OCC1 features in the effectiveness of the system was investigated and the results showed that the setting in which all the features were used led to the best performing performance. Therefore, OCC1 is used to extract emotion in the rest of the practical work in this thesis.

Chapter 6

Using Emotional Relevance to Enrich the Unified Relevance Model for Rating Prediction in Collaborative Filtering

6.1 Introduction

Recommender systems, as an example of an IR&S task (Chapter 2), attempt to alleviate users'[1] information overload by filtering documents that are not relevant to the users' interests [AT05]. Amongst these, collaborative filtering (CF) systems are the most widely used [MJZ04]. CF systems recommend an item to a user by considering data gathered from all users [Hof03]. Examples of such systems are Amazon.com[2] for products, and Netflix[3] for movies.

Chapter 4 presented a conceptual map for emotion relevance in IR both from searcher and system perspective. In order to study the practicality of the proposed map, we investigate a scenario (i.e., a branch of the proposed map) where *searchers have needs (emotional or informational) which are not explicitly identified, and where the emotion objects can be indirectly useful.* We investigate this scenario in the context of a movie recommender system.

[1] For the context of collaborative filtering, we adapt the term searcher to user.

[2] http://www.amazon.com/

[3] http://www.netflixprize.com/

Movies are one of the most popular source of entertainment and they are emotionally rich. People watch movies mainly to satisfy their emotion need. There are many theories trying to explain the strong emotional effect of movies on their audiences such as mood management, selective exposure and disposition theories, etc. (see Section 3.5.4).

Audiences of a movie are usually exposed for a longer period to the emotional content than when watching other multimedia contents (e.g., one or two hours for movies in comparison to few minutes in video clips). This longer exposure results as a stronger emotional experience in the audiences. The audiences can go through all sort of emotional experiences: they can get emotionally attached to the story and the main character, and experience a great gratification or sorrow if something good or bad happens to the main character at the end of the movie (see disposition theory in Section 3.5.4).

All of these emotional experiences and their duration make CF users quite sure about their experience of the movie. Therefore, the reviews they provide, based on their emotional experiences, are valuable sources of information and can be used to represent the movie. Therefore, we considered the emotion extracted from the movie reviews to represent movies.

In addition, movies are developed for a general audiences, their story is thus written based on a common-sense background. The same applies to their plot summaries, written to entice people to watch movies. It thus make sense to extract emotions from them. Therefore, we also considered the extracted emotion from plot summaries to represent movies.

We propose a novel approach for CF by integrating emotion information along with the semantic and rating information. This is motivated by the fact that a user likes a movie for a set of latent reasons, e.g. due to its direction style. This information can improve the CF system prediction accuracy, especially when data is sparse, i.e. when there is not enough data available (e.g. ratings) to be used by standard CF techniques. Data sparsity is a well-known problem for CF systems [AT05] and in the extreme case of the so-called *cold start* problem where, there is no rating for either new users or items, giving a prediction is impossible when not using latent semantic or emotional factors. Using semantic information has shown improvements in accuracy of CF systems [MAPJ09]. Therefore, such systems are good candidates for the investigation of the effectiveness of emotion as an additional source of information. This allows us to compare the effect of emotion features to semantic features, thus providing a better understanding of emotional features.

For this purpose, in this chapter and Chapter 7, we conduct two sets of experiments with two different models, with the aim of studying, overall, two research questions: (1) whether emotional information is useful to improve the effectiveness of the rating prediction in CF tasks, in comparison to the situations where only the rating data is used (this chapter), and (2) whether this information bring an additional (complementary) information to what exist so far for movies such as actors, director and genre (Chapter 7).

The rest of the chapter is organised as follows: Section 6.2 situate our work with respect to other CF works that consider sentiment-based features. Section 6.3 gives a background on CF systems. Section 6.4 discusses the unified CF model proposed by Wang et al. [WdVR08], on which our approach is based. Our approach is described in Section 6.5, the experiment methodology in Section 6.6, the results in Section 6.7, and finally the discussion and conclusion in Section 6.8.

6.2 Emotion in Collaborative Filtering

Sentiments extracted[4] from movie reviews have been used in CF. Those approaches are complementary to our work: using sentiment as a substitute for ratings [SPI09, PL05, GZ06, kLfClC06] or relying on sentiment expressed over movie aspects (e.g. actors in the movie are good/bad) in order to provide opinionated ratings [JWMG09] could be used likewise in our work.

In contrast to previous work[5], instead of using sentiments, we use emotions as another source of information about an item to infer ratings, and show in the experiments (see Section 6.7) that emotion information extracted from movie reviews and plot summaries help tackle data sparsity.

None of the other works exploit the richer emotion information available for collaborative filtering. In psycholinguistics and its practical applications, emotion sensing requires a significantly more

[4] The essential issue in sentiment analysis is to identify the positive (favourable) or negative (unfavourable) opinion towards the topic of a text.

[5] Emotion extracted from audio features of music [KCSL05] and low-level physical features of video clips [SDP09] have been used in CF. These approaches are complementary to our work. Due to the focus of the thesis, they are not explored further.

detailed textual analysis, as emotions are a finer-grained version of sentiment. We believe that this extra information is useful for CF.

6.3 Literature Survey on Collaborative Recommendation Approach

Collaborative recommending techniques mainly make use of users' past ratings to predict the user's level of interest for an item. Collaborative recommending systems can be divided into memory- and model-based approaches [AT05]. In a memory-based approach, recommendation is made by determining the nearest neighbours of a user and/or item, and then aggregating the ratings of these neighbours. These CF techniques have the advantage of being better adapted to users with unusual tastes, but they are impractical to use [Hof03] due to scalability issues: calculating the neighbourhood for users and/or items can be time consuming, especially in real life (i.e. commercial) datasets.

The scalability problem concerns the quantity of data, i.e., the number of users and items; many collaborative recommending algorithms fail to scale to big data sets. This issue is further stressed by the fact that in real-world systems, new items and new users appear with increasing frequency, therefore generating a huge amount of data. In addition to scalability these systems suffer from sparsity problem [AT05] which refers to situations where the ratio of unrated items to rated ones is high, therefore lacking enough information for collaborative recommending systems to base their predictions on. Data sparsity can be the consequence of many factors: (1) judging is a cognitively expensive activity [HBW92]; (2) there are unpopular or unseen items [MMN02]; and (3) in special cases, known as the *cold start* problem (i.e. new user and/or new item), no ratings are available at all. The following discusses how the two types of CF systems deal with these two challenges.

6.3.1 Memory-based Approaches

Memory-based approach can traditionally be classified either as item or user based. In user-based systems, the prediction of the rating of an item for a given user will depend upon the ratings of the *same* item by other similar users. Similarly, in item-based systems the predicted rating depends upon the ratings of other similar items by the *same* user. Both user and item-based approaches only use a restricted part of the available information, because they restrict the considered ratings either to those on the same item (user-based) or for the same user (item-based). This issue is even more important when data is sparse. A state-of-the-art approach by Wang et al. [WdVR08] proposes a *unified approach* that makes use of all the possible data, i.e. both similar items and similar users when predicting a new rating in order to alleviate the sparsity problem.

6.3.2 Model-based Approaches

Model-based techniques learn a user and/or item model from data [Hof03] and are able to scale to large datasets thereby addressing the scalability problem. Model-based approaches include regression models [SKKR01] based on user and item features , associative retrieval and graph-based models [HCZ04] or clustering of user and/or items, but the most successful ones are based on dimensionality reduction techniques since they deal better with data sparsity. These latter techniques expresses high dimensional information space in terms of low-dimension latent topics. Prominent examples include Latent semantic indexing (LSI), Principal Component Analysis (PCA), or probabilistic approaches like the probabilistic LSI (pLSI) [Hof04], nonparametric probabilistic Principal Component Analysis (NPCA) and LDA [BNJ03]. Although scalability is addressed therein, data sparsity remains a problem.

Alleviating Data Sparsity In order to alleviate data sparsity, it is necessary to resort to sources of information in addition to given sources, for example, products reviews, user demographic information as additional information to rating. This is especially important with respect to the cold start problem, as the absence of ratings therein hinders the possibility of using collaborative recommending techniques relying only on rating information.

A common way to provide additional sources is to include user/item data, in particular the representation of users and items directly, to predict ratings. For example, [PHLG00] uses these representations along with a user profile to create new ratings and then apply the standard collaborative recommender techniques on the denser data. However, the similarity between an item and a user is based therein on the actual representations of their contents, and not on any underlying/latent relations among them.

A second approach tries to use this external information by capturing the latent relationship between items and users. Mobasher [MJZ04] use a memory-based approach that takes advantage an ontology and of a latent semantic model. However, creating and maintaining the ontology is a very laborious task. Park and Chu [PC09] use regression techniques where user demographical information and item meta-data are used as features. However, the user demographic information is unreliable in general, and more importantly, users are not fully characterised by the movies they rated.

Using contextual information, as additional information, in the recommendation process by supporting multi-criteria ratings and providing more flexible types of recommendations, improves the rating prediction of recommender systems [AT05]. Contextual information refers to what is available in user and item profiles [AT01b, BS97, KRBH98, UF98], not only keywords and simple demographic [MR00, PB97] but also further details by data mining [AT01a, FP96]. This thesis investigates in particular the role of emotion as additional information in contrast to rating and semantic/contextual information.

6.4 Unified Collaborative Filtering Model

This study uses the memory-based unified model by Wang et al. [WdVR08] as base model. This model, predicts the rating of an unseen item for a given user by averaging all the ratings in the user-item rating space, weighted by their contribution. The contribution is calculated using a kernel density estimation approach whereby users (or items) close to the given user and/or item in the rating space have more influence on the outcome.

Wang et al. use a metric, based on cosine similarity to calculate distances for example, for two users u and u', and define a kernel as $2 - 2\cos(u, u')$ and similarly for items. As in most other memory-based approaches, they take advantage of user similarity and item similarity embedded in the user-item rating vector space to improve the probability estimation and to address the data sparsity problem. In addition, Wang et al. use ratings from different, but similar users, for others, but similar, items. The last point makes the approach different to standard memory-based approaches.

We now describe Wang et al. approach more precisely. Let R be the random variable that takes values between 1 and M where M is the number of rating grades, and U and I be discrete random variables over the sample space of users and items respectively. For simplicity, for a given user u, rating r, and item i, let the events $R = r$, $U = u$, and $I = i$ be respectively denoted r, u, and i. The rating predicted by the model is equal to the estimation of the expectation of the the rating r given that the user is u and the item i:

$$\mathbb{E}(r|u,i) = \sum_{r=1}^{M} rP(r|i,u) = \sum_{r=1}^{M} rP(u,i|r)P(r) / \sum_{r=1}^{M} P(u,i|r)P(r) \tag{6.1}$$

$P(r)$ is easily evaluated by the ratio of the number ratings equal to r to the total number of ratings made so far. The problem is now to evaluate the probability $P(u,i|r)$ to have a given user u and a given item i knowing that the rating is r.

In order to solve this, Wang et al. use past user/item ratings to estimate the density $P(u,i|r)$, using a standard Gaussian kernel. In order to simplify the problem, users and items are supposed to be independent given a rating, and thus, the kernel defined over users and items is decomposed into a product of two separate kernels. Defining S_r as the set of user-item couples with a rating r, we have:

$$P(u,i|r) \;=\; \frac{1}{|S_r|} \sum_{(u',i')\in S_r} K_U(r_u - r_{u'}) K_I(r_i - r_{i'})$$

In order to deal with data sparsity, bandwidth parameters that are automatically learned from data are used. The bandwidth parameters are learned automatically (given a training set of users, items and associated ratings) with a variation of the EM algorithm. Including the bandwidths h_{U-R} (for Users in a Rating space) and h_{I-R} (for Items in a Rating space) in the above formula gives:

$$P(u,i|r) = \frac{1}{|S_r|} \sum_{(u',i'\in S_r)} K(\frac{r_u - r_{u'}}{h_{U-R}}) K(\frac{r_i - r_{i'}}{h_{I-R}}) \tag{6.2}$$

The idea is that if users (or items) are in average far from each other for the same rating, then the bandwidth will be high, and it would be low in the opposite case.

In order to use a kernel, it is necessary to define a distance, and hence, to define the vector space of users and items. Wang et al. [WdVR08] followed the standard memory-based approach where the user space's dimensions correspond to the different items and where the components of the vector are the ratings he or she made[6] for a given item.

For two users u and u' the cosine similarity distance on the space is defined as $1 - \cos(u, u')$. This distance actually projects users into the space defined by the cosine distance between two users. Other metrics can be considered (e.g. the Pearson correlation coefficient), but the cosine distance has been shown to perform better [Her00]. Including the normalisation factor in formula (6.1), the final formula is:

$$\mathbb{E}(r|u,i) = \frac{\sum_{u',i' \in S} r_{u',i'} e^{-\frac{1-\cos(r_u, r_{u'})}{h_{U-R}^2}} e^{-\frac{1-\cos(r_i, r_{i'})}{h_{I-R}^2}}}{\sum_{u',i' \in S} e^{-\frac{1-\cos(r_u, r_{u'})}{h_{U-R}^2}} e^{-\frac{1-\cos(r_i, r_{i'})}{h_{I-R}^2}}} \tag{6.3}$$

Experimental results show that the unified model performed better than user and item-based models, mainly since data sparsity is addressed by taking more data into account than simple item or user based models.

6.5 Approach

A number of questions remain unanswered by the rating representation, for example "*why does the user like this item?*", and "*What are the underlying interests of two users that make them similarly rate some items?*".

Our approach addresses these questions by considering emotion and semantic spaces. Each item – whether a movie, music, book, product, game, or otherwise – can be characterised by emotion and

[6] a rating of 0 is usually used when the user did not rate the item.

semantic traits. For example, the genre of a movie can be chosen among different types as comedy and drama. To compute the similarity between two items or users, we do not only consider the rating space, but also semantic spaces. These spaces need to be defined in relation to the type of items we are dealing with (e.g. movies, books, etc.).

In the case of movies, people ratings are not only influenced by the genre, but also by actors, or directors. Each of these semantic spaces (i.e. actor, genre, director) contain valuable information that can be used as sources of information for estimating whether an item should be recommended to the user or not.

Regarding emotions, this thesis asserts that People ratings are also dependent of emotions expressed in movie reviews and plot summaries. We here use an *emotion space* to explorer his hypothesis.

Our goal was to define new vector spaces based on emotion and semantic information and calculate user similarity and item similarity in these spaces. The following explains the method of constructing emotion and semantic spaces and the calculation of and the cosine similarity between two users/items.

6.5.1 Construction of a semantic space

In order to construct a semantic space for the genre of a movie, an initial intuition is to build two vector spaces for each semantic space (one for the users, one for the items) and define the representation of users and items, so to be able to compute distances in those spaces.

Let us illustrate the process with the genre space. The different genres a movie belongs to can be extracted from IMDb website, e.g., comedy, horror, drama. A movie i, is represented as a vector $g_i^{'} = (0, \ldots, 0, 1, 0, 0, 1, \ldots, 1, \ldots)$ where each dimension corresponds to a genre, 1 corresponds to the movie belonging to this genre, and 0 otherwise; a movie can belong to several different genres. From this representation, we construct the user vector by giving more importance to the genres of the movies the user liked, i.e. those for which the rating is high. More formally, the vector of a user in the genre space is defined as $g_u^{'} = \sum r_{u,i} g_i^{'}$ where $r_{u,i}$ is the rating of the item i for user u.

In the above defined genre space the cosine metric seems rational to employ, however, there are issues related to data sparsity. Consider the case where users will judge similar items that belong to comedy and to comedy-drama. That is, a given user will either like both genres or dislike both. If this is the case, then considering the defined genre space will not capture this information since two items belonging to comedy and comedy-drama respectively will have a similarity of 0. This may not be problematic with genres as there are not many of them, but it might be more significant, for example, with actors or directors who are much more numerous, and for which data sparsity might have important effects.

To capture the interpendance between different genres, we chose to represent users and items in a similarity space. Users are represented as vectors in a space where each dimension corresponds to an item i and the vector component is the similarity between item i and the user u; $g_u = \left(\ldots, \cos\left(g_u^{'}, g_i^{'}\right), \ldots\right)$, is similarly defined.

To illustrate the latter step of the space construction, consider the matrix on in Figure 6.1. Rows correspond to users, items to column, and cells to the cosine similarity between a user and an item in the space of $g^{'}$. We use rows as the vector representation of users and columns for items, in order to compute cosine similarity in the density estimation formulas. Given the cosine distance between two users or two items, we then construct a distance based on it: the expression $2 - 2\cos(g_u, g_{u'})$ for users and $2 - 2\cos(g_i, g_{i'})$ for items – similar to Wang et al. [WdVR08].

This method solves the interdependancy problem, i.e. the *comedy versus comedy-drama problem*: A comedy movie will be similar to the comedy-drama movie in the newly defined space since *the comedy and the comedy-drama movies are rated similarly by different users*. Symmetrically, two users will be similar in the genre space if they rate similarly the same genres (and not the same items).

6.5.2 Construction of an emotion space

OCC1 categorises emotion into 22 emotional categories as defined by the OCC emotion model (see Section 5.4). A movie i in an emotion space (i.e., review or plot summary) is represented as a vector $p_i^{'} = (1, \ldots, 0, 0, 1, \ldots)$ where each dimension corresponds to an emotion extracted from its

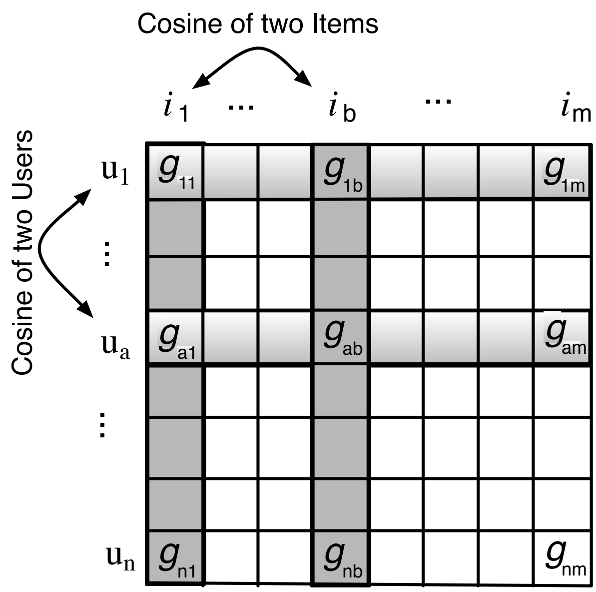


Figure 6.1: User-Item Genre Matrix

reviews. A component is set to 1 if the corresponding emotion was detected and 0 otherwise. Given the movie representation we construct the vector representing a user by giving more importance to emotions present in the movies the user liked, i.e. those for which the rating is high. More formally, the vector of a user in the emotion space is defined as $p_u^{'} = \sum r_{u,i} p_i^{'}$ where $r_{u,i}$ is the rating of the item i for user u. Using the cosine to compute the similarity between two item or user vectors seem natural, as it the case of the rating space.

Finally, we create a user-item emotion space where users are represented as vectors in the space and each dimension corresponds to an item i, and the vector component is the similarity between item i and the user u. The vector can be rewritten for such a user as $p_u = \left(\ldots, \cos\left(p_u^{'}, p_i^{'}\right), \ldots\right)$, and in a similar fashion for items. Employing a matrix representation, the rows correspond to the vector representation of users, and columns as the vector representation of items. The cosine similarity used in the density estimation formulas can then be used to construct a distance based on it, defined as $2 - 2\cos(p_u, p_{u'})$ for users and $2 - 2\cos(p_i, p_{i'})$ for items.

6.5.3 Integrating emotion and semantic spaces in the unified approach

The following uses the emotion and semantic information discussed in the prior section and considers combining emotion, semantic and rating spaces using the Kernel Multiplication method. We here only show formulas using the rating and genre spaces, but they can straight forwardly be extended to use all the above defined emotion and semantic spaces.

Kernel Multiplication

To estimate $P(u, i|r)$ in Formula 6.1 we consider the rating and various emotion and semantic spaces, and equate $P(u, i|r)$ to $P(r_u, r_i, g_u, g_i|r)$. A cosine based kernel density estimator is defined (like in Wang et al. [WdVR08] approach). The final rating prediction $\mathbb{E}(r|u, i)$ is defined by:

$$\frac{\sum_{u',i' \in S} r_{u',i'} e^{-\frac{1-\cos(r_u, r_{u'})}{h^2_{U-R}}} e^{-\frac{1-\cos(r_i, r_{i'})}{h^2_{I-R}}} e^{-\frac{1-\cos(g_u, g_{u'})}{h^2_{U-G}}} e^{-\frac{1-\cos(g_i, g_{i'})}{h^2_{I-G}}}}{\sum_{u',i' \in S} e^{-\frac{1-\cos(r_u, r_{u'})}{h^2_{U-R}}} e^{-\frac{1-\cos(r_i, r_{i'})}{h^2_{I-R}}} e^{-\frac{1-\cos(g_u, g_{u'})}{h^2_{U-G}}} e^{-\frac{1-\cos(g_i, g_{i'})}{h^2_{I-G}}}} \tag{6.4}$$

6.6 Experiment

6.6.1 Test Collection

Evaluation was done using the Movielens dataset, which was created by Grouplens through the Movielens web site [RIS+94]. An evaluation was performed using two MovieLens datasets containing 100,000 ratings for 1682 movies by 943 users (100K dataset) and 500,000 ratings for 3900 movies by 3020 users (500K dataset) respectively. The latter was extracted from the dataset containing 1 million ratings for 3900 movies from 6040 users (1M dataset) by randomly selecting half of the users. We did not use the 1M dataset for computational reason. In each of these datasets, there are at least 20 movie ratings per user on a scale takes values from 1 (not liked) to 5 (most liked).

6.6.2 Source of Semantic and Emotion Information

We extracted the information needed to define the different semantic and emotion spaces from the IMDb website[7]. We considered genres, actors, and directors as our semantic spaces. Emotion extracted from plot summaries and movie reviews was used to define our emotion spaces.

6.6.3 Metrics

We evaluated our approach and hypotheses using three standard collaborative filtering metrics. First, two of the most widely used metrics of collaborative filtering, MSE (Mean Squared Error) and MAE (Mean Average Error), are used to measure the average error made at the rating level. MAE is the average absolute deviation between observed and predicted ratings, and MSE is the average squared deviation. MSE will penalise systems for having a comparatively small number of big deviations rather than those having a big number of small deviations. For both measures a smaller value indicates better performance. In the formulas for MAE and MSE, N is the number of test ratings, $r_{u,i}$ the actual rating, and $\widehat{r}_{u,i}$ the estimated rating:

$$MAE = 1/N \times \sum_{u,i} |r_{u,i} - \widehat{r}_{u,i}| \qquad MSE = 1/N \times \sum_{u,i} |r_{u,i} - \widehat{r}_{u,i}|^2 \tag{6.5}$$

The goal of CF is often to return the relevant items to the user, such as the top rated movies. In this case, the performance of a CF algorithm with respect to those movies is better measured using mean average precision (MAP).

6.6.4 Evaluation Protocols

Evaluation Procedure

10-fold cross validation was performed with user ratings randomly split into two sets of equal size, one for observed items and the other for held-out items. Held-out items for a user were discarded

[7]The Internet Movie Database (IMDb, http://www.imdb.com/) is an online database of information related to movies, actors, television shows, production, etc.

for prediction (i.e., density estimation), and were only used for evaluation purposes, i.e. to compare the predicted rating with the observed one.

The spaces were constructed as explained in Section 6.5.1 and 6.5.2 and for each space, the user and item bandwidths were calculated using the EM algorithm as described in [WdVR08]. We then evaluated the performances of the model for test set users: The ratings for the held-out items were predicted and compared to their real value. To ensure result consistency, we repeated the experiments twenty times for each model, each time with a different training/testing split.

Complexity

The computational complexities derived from way of creating the distance matrix, estimating the bandwidth parameter with the EM algorithm and predicting the rating are $O(AB^2 + B^2A)$, $O(ABN^2)$ and $O(kN^2)$ respectively where A is the number of items, B the number of users, k the number of spaces and N the the top-N nearest neighbour[8]. In order to decrease the complexity of calculation (Formula 6.4), the top-50 nearest users were considered for the test user and the top-50 nearest items for the test item. While it would have been theoretically more optimal to use the nearest items/users with respect to the kernel K, for complexity reasons, the distance we used was the distance in the rating space as it gave us good results.

Evaluation Methodology

Five experiments were conducted to examine the effect of one space on rating prediction performance. We modified the Wang et al. [WdVR08] by including two new kernels for items and users, based on one of the following spaces: Rating (R), Genre (G), Actor (A), Director (D), Plot Summary Emotion (P), and Movie Review Emotion (M) spaces. The results are presented in the Tables 6.1 and 6.2. Pairwise combinations of rating were used in addition to other spaces for predicting ratings. The purpose is to isolate the potential contribution of each space to the rating. In each pair a total kernel was defined as the product of the two individual kernels. Each scenario is labeled by the letters corresponding to the spaces, for example, RM refers to the combination of rating, and movie review emotions spaces.

[8]This was derived from estimations given in [WdVR08].

Table 6.1: MAE for 100K and 500K test collections

	Baseline	RA	RD	RG	RP	RM
100K	0.911	0.859 +5.7%	0.851 +6.5%	0.859 +5.7%	0.879 +3.5%	0.875 +3.9%
500K	0.881	0.82 +6.9%	0.804 +8.7%	0.814 +7.6%	0.84 +4.6%	0.828 +6%

Table 6.2: MAP for 100K and 500K test collections

	Baseline	RA	RD	RG	RP	RM
100K	0.654	0.729 +11.4%	0.731 +11.7%	0.704 +7.6%	0.69 +5.5%	0.697 +6.5%
500K	0.705	0.778 +10.3%	0.776 +10%	0.76 +7.8%	0.75 +6.3%	0.76 +7.8%

6.7 Results

The results show that emotional features consistently play a role in improving the recommendation quality in comparison to the rating only space (Baseline). Emotion extracted from the movie plot summary and movie review emotion spaces affect system performance differently. Movie review emotion space performs better than movie plot summary emotion space. This is perhaps because movie reviews include a richer emotional content than plot summary does.

Among semantic spaces, in terms of the number of features, genre space is the closest one to emotion spaces. Movie review emotion and genre spaces have the same improvement in terms of MAP for the 500K dataset. This situation has not been observed for 100K dataset. This suggests that the larger the dataset, the better the emotion spaces perform. This motivates to further investigate the role of emotion in a larger dataset, and for this purpose, a model-based CF approach is required (Chapter 7).

The behaviour of emotion spaces with respect to the size of datasets are different from actor and director spaces. As the data size increases, consistent improvements in MAP are obtained when using emotion spaces, whereas this pattern does not hold for actor and director spaces.

Although in general using emotion spaces results in a lower performance than using semantic spaces (apart from the MAP for 500K dataset in genre space), the semantic spaces are constructed based on manually created metadata, i.e. human intelligence, which is very costly and time consuming whereas emotion features were automatically generated by the emotion extraction system. This makes emotional features a good alternative to manually created semantic features.

Although the improvements achieved by emotion spaces are not statistically significant (similar to semantic spaces), the comparative performance of emotion spaces with semantic spaces makes the results encouraging. Previous studies [MJZ04] has shown the usefulness of semantic information in improving the accuracy of CF systems. Therefore, a more sophisticated model is required to investigate the usefulness of such spaces (Chapter 7).

6.8 Chapter Summary

This chapter investigated the effect of emotion features in improving the performance of a memory-based CF system. A modification of Wang et al. [WdVR08] CF model that takes into account semantic or emotion spaces was proposed. We proposed to combine semantic and emotion information with that of the ratings, by including emotion and semantic information directly within the density estimator.

The experiments showed that using emotional features significantly improves the rating prediction, and that semantic subspaces (director, genre, actor and rating) lead to significantly better results over the baseline (using only ratings). Emotion extracted from the movie plot summary and movie review emotion spaces affect system performance differently. We believe that this is due to the richer emotional content in opinionated movie reviews compared to the relatively more emotionless plot summary texts.

This study indicates that the improvement of emotion spaces increases with the size of the datasets. This motivated the study of the role of emotion with larger datasets since memory-based approaches do not scale (well) (see Section 6.3). On the other hand, preliminary results for combining emotion and semantic features using modified Wang et al. [WdVR08] model, did not show improvements over the scenarios where only semantic or emotion spaces were employed individually, indicating the possibility that the model is not sophisticated enough to take advantage of the information from different spaces. Thus, to further study the role of emotion features. In the next chapter, we propose to use emotion and semantic spaces in a model-based CF approach based on LDA and boosted trees.

Chapter 7

Handling Data Sparsity in Collaborative Filtering using Emotion and Semantic Based Features

7.1 Introduction

The results of the study in Chapter 6 indicate that incorporating emotion features improves the accuracy of rating prediction in CF. However, the proposed approach suffered from the following limitations:

1. the similarity between items and users is calculated based on the direct representations of the features in emotion and semantic spaces, and do not consider the underlying/latent relations among them;
2. the scalability problem associated with memory-based approaches do not allow us to study the role of emotion in large datasets where the improvement achieved by using emotion spaces in CF systems is expected to increase;
3. the preliminary study shows that a more sophisticated model is required to combine the information encapsulated in different spaces; and

4. Wang et al. [WdVR08] model relies on past item ratings in order to construct the emotion spaces and therefore cannot predict ratings for the items that do not provide such information (e.g. new items). Therefore, it is not possible to investigate the role of emotion in item cold start problem.

In this section, building upon the promising results obtained using emotion in a memory-based approach to CF, we propose to use emotion and semantic spaces in a model-based approach that allows to address these limitations.

From a technical perspective, for each semantic (e.g. actor) and emotion (e.g. plot summary emotion) space we propose to construct latent groups of users. In order to do so, we extend a well-known model-based approach, namely Latent Dirichlet Allocation (LDA) [BNJ03]. In each space, we propose a methodology to compute the probability that a given user likes an item. Finally, in order to predict a rating, the information about the different spaces is aggregated using standard machine learning techniques.

From an experimental point of view, we have conducted extensive experiments where we vary the sparsity of the dataset and compare our models to two state-of-the-art CF approaches. Furthermore we present preliminary experiments in an item cold start scenario and analyse qualitatively the latent spaces uncovered by our extended LDA approach.

The rest of the chapter is organised as follows: our approach is described in Section 7.2, experiment methodology in Section 7.3, results in Section 7.4, and finally the discussion and conclusion in Section 7.5.

7.2 Approach

In this section, we describe a methodology for extending LDA to take into account item semantic and emotion information. We first suppose that a movie can be described by a set of features in a space s. For example, in the actor space, a movie is described by the set of actors that played in the movie. The different spaces considered are detailed in Section 7.2.1.

For a given space s, we first evaluate the probability that a user u, defined by his/her past ratings, likes the movie m that is described by a feature f (e.g. De Niro played in the movie):

$$P(+|f, u, s) \tag{7.1}$$

where $+$ denotes the event that a user likes a movie. The LDA approach is used to compute this value, in Section 7.2.2.

As a movie is described by a set of features (e.g. actors in the actor space) it is therefore necessary to aggregate the probability in Eq. (7.1) over these features, i.e. to estimate the probability that a movie m is liked by user u:

$$P(+|m, u, s) \tag{7.2}$$

This probability can be estimated given different probabilities calculated with Eq. (7.1) for each possible feature where the presence of a feature independently influences the relevance of an item (detailed in Section 7.2.3).

The underlying characteristics of feature spaces vary, they provide different views on the same subject. Since there is no information about how to best to combine these features, we chose to employ a machine technique to combine the predictions made on individual spaces (Section 7.3.3). The performance of each individual feature space, and discussion of whether they complement each other is in Section 7.4.

7.2.1 Feature Spaces

Three types of feature spaces are considered. First the movie space with the movie itself as a feature, where, the probability of a movie to be liked is directly given by Eq. (7.1) where f is the movie at hand. It is the space on which most of the CF systems are based since it relates movie and users by their ratings.

Second, the semantic features spaces of which there are three, actor, director and genre spaces[1]. A movie is represented by the set of features that characterise it in each of these spaces, e.g., in the actor space, the movie *Dr. Mabuse: The Gambler* is associated with the list of its actors, i.e. Rudolf Klein-Rogge, Aud Egede Nissen, etc.

[1] As they are readily available on the Web and are likely to be good predictors for user ratings.

Finally the last type of feature space is the emotion space where for each movie, emotion features are constructed based on the emotion extracted from its reviews or plot summaries. Emotions were extracted by OCC1 (see Section 5.4).

Movie reviews and plot summaries are texts composed by a set of sentences. OCC1 makes a binary decision about each emotion for a sentence, i.e. decides whether the emotion is present or not. Let T denotes the set of texts associated to a movie (either reviews or plot summaries), and S_t the set of sentences associated to a text $t \in T$. For each sentence k of a text t, we can use this classifier to construct a vector of 24 components, each of those associated with one of the 22 emotions and two cognitive states (see Section 3.3.2). Each component can take the value 0 (the emotion is not present in the sentence) or 1 (the emotion is present).

In order to represent the emotions in a text $t \in T$, we sum the emotion vectors of the sentences in t, and normalise the values by dividing by the number of sentences. Then, for a set of texts T we sum the vectors corresponding to the individual texts and normalise again, this time by the number of texts. Formally an emotion vector for a set of texts T is defined as

$$emotions(T) = \frac{1}{|T|} \sum_{t \in T} \frac{1}{|S_t|} \sum_{k \in S_t} emotions(k) \tag{7.3}$$

The LDA model (see Section 7.2.2) needs discrete components for the emotion vector and as the distribution of values for each emotion can be very different we used a non-parametric way of discretizing. Each value was assigned its corresponding quartile, e.g., if the values of the component corresponding to the emotion *fear* are distributed evenly in the four quartiles $[0, 0.3)$, $[0.3, 0.4)$, $[0.4, 0.75)$ and $[0.75, 1]$, then a value of 0.32 would be transformed into 2. This would in turn be represented by *fear-2* by LDA.

7.2.2 Building Latent Spaces

This section details the method to estimate the probability that a movie is liked or disliked because of a feature. We build upon LDA [BNJ03] which is a generative probabilistic model for discrete data collection used mainly for textual corpora.

LDA represents documents as a probability distribution over latent topics, where each latent topic is a distribution over words. Documents that have similar topics should share the same latent topic distribution. This corresponds to CF where users who share the same ratings for the same items have related interests, and should thus be in the same *latent groups* defined by a similar distribution over features.

In the LDA approach to CF described[2] of [BNJ03], a user is defined by a probability distribution over a set of latent groups. Each group in turn defines a probability distribution over the movies that are liked by the users represented by this group. It is then possible to compute the probability that a user likes a movie by marginalising over the different possible latent groups. More formally, LDA defines the probability to observe a series of movies $\mathcal{M}_u^+ = (m_1, \ldots, m_n)$ liked by a user u:

$$p\left(\mathcal{M}_u^+ \middle| \alpha, \beta\right) = \int p\left(\theta|\alpha\right) \left(\prod_{i=1}^{n} \sum_{z=1}^{T} p\left(z|\theta\right) p\left(m_i|\beta_z\right)\right) \mathrm{d}\theta \tag{7.4}$$

where T is the number of latent groups, θ follows Dirichlet distribution of hyper-parameters α, z (the latent group) follows a multinomial distribution given by θ and finally the probability of liking a movie m given the latent set of users z follows a multinomial distribution given by β_z. The model is fully specified by the α and the β_z for each possible latent group z. Those hyper-parameters are learnt by maximising the likelihood over the dataset (Section 7.3.3).

However, one of the problems of LDA is that it gives high probabilities to popular movies. Consider two movies judged the same number of times, the probability given by LDA will rank the two movies in order of their probability to be liked. However, if the first movie has been judged by all the users and liked half of the time, it will have the same probability as another movie judged by only half of the users but liked all the time. This is an undesirable artefact of the model caused by the fact that LDA assigns high probabilities to movies (in the movie space) that are liked *and* judged by many users. Thus, the probability assigned to a movie m for a given latent group does not correspond to the probability that this movie would be liked by a user of this latent group, but rather to the probability that if we pick at random a movie liked by a user of this latent group, it will be m. Formally, LDA gives us the joint probability $P(+, m|u, s)$ instead of the conditional probability $P(+|m, u, s)$. This *popularity* problem becomes even worse when using feature spaces since some features are present in just a few movies (like an actor who has played only in a few not so popular movies).

[2] To ease the reading, we adapted here the notations and concepts. For example, we refer to latent topics as *latent groups* since we are performing LDA on non textual information.

We propose to alleviate this problem by considering negative information (i.e in movie space, movies that have been disliked, or in semantic or emotion space, features that appear in a movie that has been disliked). That is, we define LDA as a generative process of a series of couples feature-decision $\mathcal{F}_{u,s} = ((f_1, d_1), \ldots, (f_n, d_n))$ where f_i is a feature and d_i its associated decision, either *liked* (+) or *disliked* (-):

$$p(\mathcal{F}_{u,s}|\alpha, \beta) = \int p(\theta|\alpha) \left(\prod_{i=1}^{n} \sum_{z=1}^{T} p(z|\theta)\, p(f_i, d_i|\beta_z) \right) \mathrm{d}\theta$$

Let us illustrate how the set of couples is computed in the actor space. Assume that a user has (i) liked a movie with actors a, b and c; (ii) disliked a movie with actors a and b; (iii) liked a movie with actors a and d. This user would be represented by the couples $(a, +)$, $(b, +)$, $(c, +)$, $(a, -)$, $(b, -)$, $(a, +)$ and $(d, +)$.

Besides addressing the popularity problem, this approach also has two advantages. First, we consider more information to learn the LDA latent groups, since negative information is used. Second, user groups reflect not only features (e.g. actors) that appear in the movies they like, but also in the movies they don't like, thus providing richer information.

The LDA model is used to compute the posterior distribution of whether the feature f indicates a movie liked (+) or disliked (-) given the past user interaction $\mathcal{F}_{u,s}$ and the learnt parameters α and β, that is

$$P(\pm, f|u, s) = p(\pm, f|\mathcal{F}_{u,s}, s, \alpha, \beta)$$

In the next section, this probability is used to derive the final formula corresponding to a movie being liked in a feature space.

7.2.3 Probability Estimation based on a Feature Space

This section presents our methodology for calculating the probability that a movie is liked given a user and corresponding movie features. Our approach is based on a simple *averaging* method where the probability that the movie is liked is the expectation that the movie is liked due to each of its features[3]. The probability $P(+|m,u,s)$ that user u likes movie m in the feature space s is

$$\begin{aligned} P(+|m,u,s) &= \sum_{f \in F} P(+,f|m,u,s) \\ &= \sum_{f \in F} P(+|f,u,s)P(f|m,s) \end{aligned} \tag{7.5}$$

where F is the set of possible features for a given movie and where we assumed that (i) features are examined one at a time to make a decision about whether a movie is liked or not. In this case, f and f' are disjoint events whenever $f \neq f'$; (ii) when the feature is known the judgment does not depend any more on the movie, i.e. $P(+|f,m,u,s) = P(+|f,u,s)$; and (iii) a movie that has a given feature is independent from the user, i.e. $P(f|m,u,s) = P(f|m,s)$.

Eq. (7.5) reduces to the estimation of two quantities: the probability of considering the feature f given a movie m and a space s, i.e. $P(f|m,s)$, and the probability that a user u likes a movie given that it has the feature f, i.e. $P(+|f,u,s)$. The latter probability is straightforward to estimate, since the probability $P(+|f,u,s)$ can be rewritten as

$$P(+|f,u,s) \quad = \quad \frac{P(+,f|u,s)}{P(+,f|u,s) + P(-,f|u,s)} \tag{7.6}$$

where $P(\pm,f|u,s)$ is the probability that the feature f occurs in a movie that is liked (or disliked) by the user u in the space s, which is given by our extended LDA.

Note that when only a few observations are available for a given movie the estimations given by Eq. (7.6) can be unreliable, this is especially true when data is sparse. We tried different smoothing techniques, the best performing one was the Laplace smoothing:

$$P(+|f,u,s) \quad = \quad \frac{P(+,f|u,s) + \epsilon}{P(+,f|u,s) + P(-,f|u,s) + 2\epsilon} \tag{7.7}$$

The ϵ value is set to $0.001 \times |s|^{-1}$ where $|s|$ is the number of features of space s. This scaling was necessary in order to adapt to the different spaces where the number of features can vary greatly.

Unless we have an a priori reason to give more importance to a feature (e.g. to give a higher importance to the main actors), we can assume a uniform distribution over the features present in movie m in space s of $P(f|m,s)$. Let $F(m,s)$ be the set of features present in movie m in the

[3] Other forms of aggregation were tried but preliminary results suggested that they would not improve over this simple method.

space s and $\#F(m, s)$ the set cardinality, the probability $P(f|m, s)$ is $\frac{1}{\#F(m,s)}$ if f is a feature of space s for the movie m and 0 otherwise. Putting the derived quantities back into Eq. (7.5), the final prediction formula is

$$P(+|m, u, s) = \sum_{f \in F(m,s)} \frac{1}{\#F(m, s)} \frac{P(+, f|u, s) + \epsilon}{P(f|u, s) + 2\epsilon} \tag{7.8}$$

Note that in the case of the movie space, each movie is defined by one distinct feature and the sum reduces to one term. Finally, to compute the final rating prediction for a given item we combine the information from the different spaces as given by Eq. (7.8), using boosted trees (see Section 7.3.3).

7.3 Experiments

7.3.1 Test Collection

Our approach is evaluated on the *100K* and *1M* MovieLens datasets presented in Section 6.6.1. As discussed in Section 6.6.2, we extracted the information needed to define the different semantic and emotion spaces.

Metrics To measure the performance of the models, we used three metrics presented in Section 6.6.3: MSE (Mean Squared Error), MAE (Mean Average Error), and MAP (Mean Average Precision). Results from MSE and MAE were similar, and therefore, we only report the former. For clarity, we include the complete results and report the cases where MAP had a different behaviour than MSE.

7.3.2 Evaluation Protocols

The variability of our results was investigated by performing a 10-fold cross validation where 70% of the users were used each time to train the LDA (Section 7.3.3), and 20% to identify the number of latent groups for LDA and to train the boosted trees based on LDA output (Section 7.3.3). The remaining 10% were used for performing the test.

In order to study the impact of sparsity on our models, following standard methodology, we randomly removed some ratings from the training set so that the maximum number of rated items per user is below a given threshold (10, 20 and no limit, coined *full*), where 10 represents the highest sparsity and full the lowest.

The last processing step divides, for each user, the set of rated items into two. One set is used to represent the past history of the user, i.e. to compute the user representation in the various feature latent spaces. The second set of items are held out, and their predicted rating is computed with each model before being compared to the real value in order to measure the performance of the model. The following two splitting methodologies were considered:

Random For each user, we randomly divide the items in two. In doing so, some users might have rated an item that is held out for testing for another user.

Cold Start 10% of the items that have been rated by the test users were randomly selected to be the held out set for *all* users. In order to ensure that it is a cold start, we also removed the ratings of these items in the whole training set.

Evaluation Methodology We tested the performance of our models with different combinations of the features spaces, i.e. Movie (M), Genre (G), Actor (A), Director (D), Review Emotion (R) and Plot Summary Emotion (P) spaces. The configuration M is similar to LDA, but as explained in Section 7.2.2, does make use of negative information.

The first set of experiments investigate the effect of individual space, the models are represented by the initial Letter associated to each space (M, G, A, D, R or P). In the second set of experiments, we investigated the effect of a combination of spaces by using, besides the movie space, only emotion spaces (MPR), semantic spaces (MGAD) and all the spaces (MGADPR).

Finally, we used three different baselines. First, as a threshold, we report the performance of a constant rating estimator that returns the mean of the ratings in the training set. Secondly, we report the performance of the original LDA approach (identified as LDA) along with our model on movie space (identified as M) for comparison. Third, we report the performance of nonparametric probabilistic principal component analysis (NPCA) presented by Yu et al. [YZLG09], which has been shown to outperform other state-of-the-art approaches in the literature and thus is a stronger baseline.

In the cold start situation, systems that rely on past item ratings cannot predict ratings for the items that do not provide such information (e.g. new items). This means that NPCA, original LDA, or our model based on movie space (M) cannot be employed to address the cold start problem. In cases where M is combined with others, we simply removed the space M, leading to combinations based on emotions (PR), semantic spaces (GAD) and all the spaces (GADPR). It can be argued that the review emotion space (R) should not be used when we are dealing with the cold start problem. However, in our experiments, we consider the reviews as a movie feature rather than a user feature since any individual who is not part of the CF system can give these reviews. Moreover, we are interested to see the effect of utilising the review emotion space when there is no rating available for a movie.

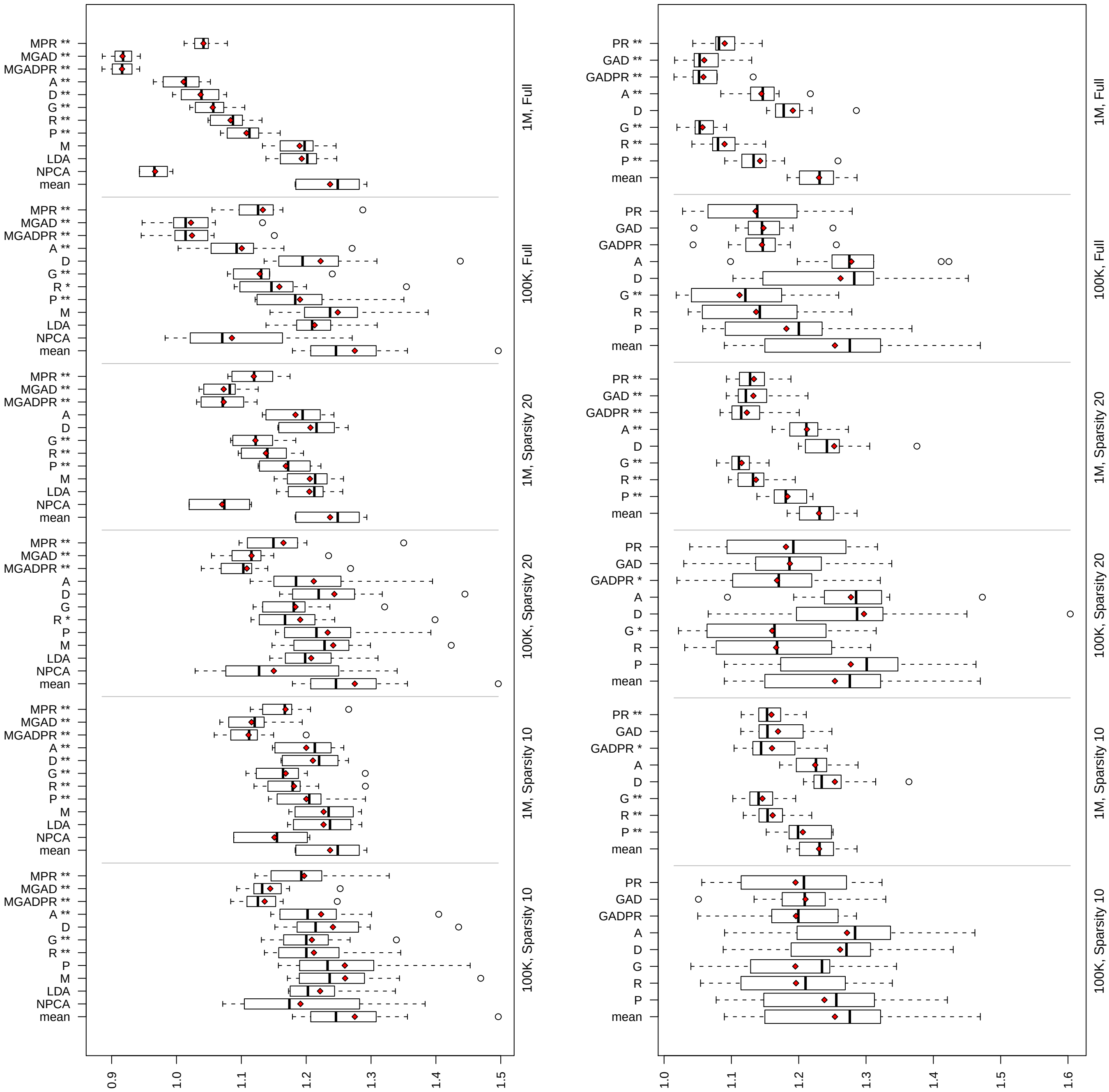
7.3.3 Optimising Parameters

LDA

Binary relevance judgments are required to train the LDA model. In addition, two sets of hyper-parameters, the number of latent groups T and the initial α and β are also required. In transforming to binary values, ratings of 3 (neutral) were discarded 1-2 were mapped to negative, and 4-5 to positive.

The initial values of α and β were set according to Misra et al. [MCY08]. The number of latent groups greatly influences the performance of the LDA approach. The appropriate number of latent groups were found by standard methods from dimensionality reduction techniques based on the likelihood over a held out set of training data [MCY08].

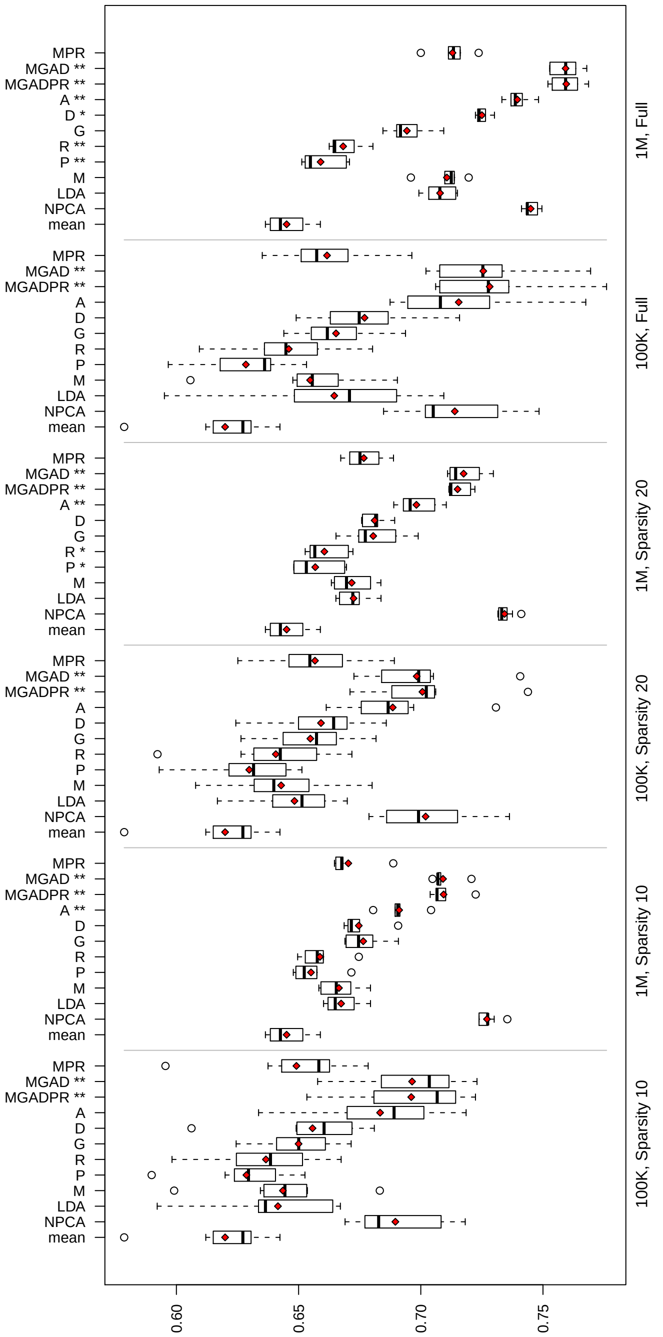
For each space and dataset, several different quantities of latent groups were tried: 3, 5, 10, 20, 35, 50, 100, 120, and 150. The maximum number of latent groups was set to 150 for computational



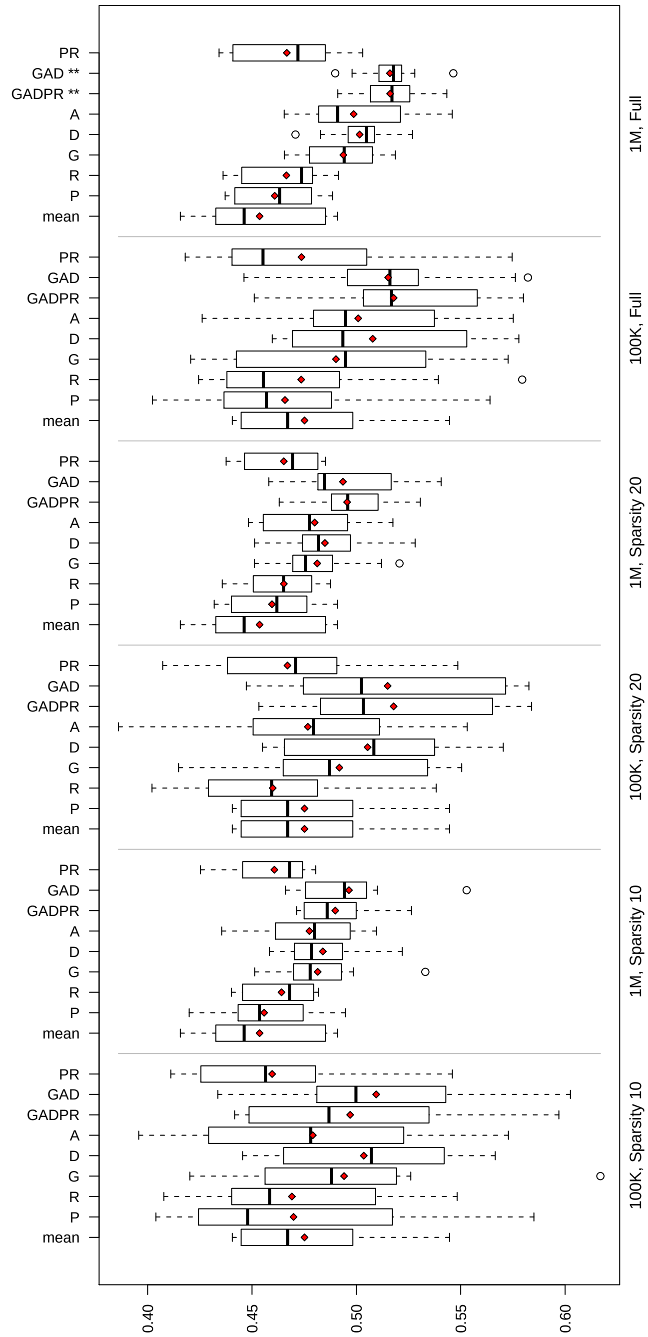
(a) Random split

(b) Cold start split

Figure 7.1: Box plot of MSE for different sparsities (no sparsity, sparsity=20, sparsity=10) and datasets (100K, 1M). The lower the value, the better the performance. The diamond represents the mean of each model.



(a) Random split



(b) Cold start split

Figure 7.2: Box plot of MAP for different sparsities (no sparsity, sparsity=20, sparsity=10) and datasets (100K, 1M). The higher the value, the better the performance. The diamond represents the mean of each model.

reasons. The number of latent groups was selected by maximising the likelihood of observations over the second subset of the training set.

Boosted Tree

The final rating was predicted by the standard machine learning technique of gradient boosted trees [Fri01]. The features given to the boosted trees are the set of probabilities governed by Eq. (7.8), one for each of the different spaces used in a given model, the output is a predicted rating between 1 and 5.

Note that even in the case of single space-based models, e.g. M, G or A, boosted trees are still useful since they map the probability of a movie to be liked to the rating scale. In order to ensure a fair comparison, we also used boosted trees to predict a final rating for the LDA-based model.

The parameters used for our experiment were found during preliminary experiments. The maximum number of trees were set to 2000, the maximum tree depth to 3 and a gaussian cost function for optimising the MSE. We used 65% of the data to train the boosted trees leaving 35% to control for over-fitting.

7.4 Results & Discussion

In Section 7.4.1 we analyse the results for models based on individual and multiple spaces, and discuss the latent groups discovered by LDA in Section 7.4.2.

7.4.1 Quantitative Study of the Performance of the Models

Figure 7.1a (random split) and 7.1b (cold start split) show the box plots for the MSE measure, for the two test collections (100K and 1M) and for different levels of sparsity (10, 20, and full). Each box plot reports, over the 10 cross validation sets, five important pieces of information namely the minimum, first, second (median), third, and maximum quartiles[4]. A paired t-test was done between

[4]Further information can be found in [MTL78].

measures obtained for each user to check the significance of the difference with the baseline (M in Figure 7.1a and mean in Figure 7.1b). Symbols (*) and (**) denote that the corresponding model had results different from that of the baseline in all the cross validation sets with the confidence levels $(p < 0.05)$ and $(p < 0.01)$ respectively.

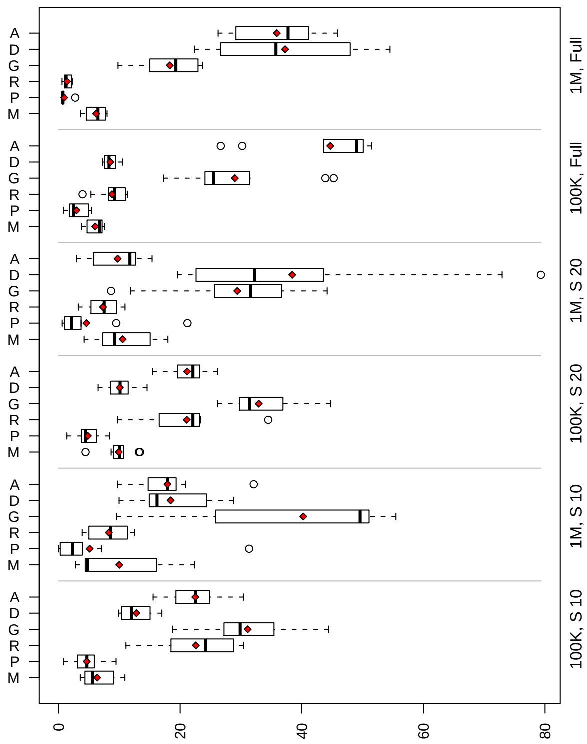


Figure 7.3: Box plot of the importance of each space in MGADPR model calculated by boosted trees for different sparsities (no sparsity, sparsity=10) , and datasets (100K, 1M). The higher the value, the higher the importance. The diamond represents the mean of each model.

Main results For models based on several feature spaces, the proposed model combining all spaces (MGADPR) consistently and significantly outperformed other models, including NPCA (shown in Figure 7.1a). With low sparsities, it has a better median and similar variance, and with high sparsities, it has a slightly better median but a much lower variance. This shows that substantial performance improvements can be made by integrating multiple sources of information for predicting ratings.

The model using movie space (M) has a slightly better (statistically significant) performance than original LDA (identified as LDA in Figure 7.1a) with complete data (1M, Full). However, with high sparsity, this model performed poorly. Hence for this space, popularity is important when data is sparse. The same pattern was not observed for the models using semantic or emotion space. Preliminary experiments (not reported here) showed that the models based on our extended LDA approach (with negative and positive information) did outperform those based on the original LDA model. The improvement of our model, with semantic or emotion spaces, is significantly better than original LDA model.

Cold start The state-of-the-art methods discussed in this chapter do not handle the cold start problem and we therefore consider the mean of ratings as our baseline. Figure 7.1b shows how our proposed approach performs in the case of the cold start problem. There is an overall increase of 0.1 in MSE for GADRP over the baseline, and this holds for different level of sparsity. The comparison between the model using emotion spaces (i.e PR) and the one using semantic spaces (i.e. GAD) shows a comparable performance between them (sparsity *full* and 20), but with high sparsity (10), emotion space based models are performing better.

MAP and MSE By comparing the MAP (not shown here) and MSE performances, we can distinguish systems that are good at predicting the rating on average from those that improve the top ranking. Movie and actor feature spaces were observed to perform well at predicting top ranked items (at the price of a higher variability in the 100K dataset), however, models based on emotions and to some extent genre (G) did not perform well in terms of MAP compared to the other spaces. Thus, while they are useful for accurate rating prediction, they are not good predictors for highly rated items.

Analysing Feature Spaces Let us now analyse the performance of each feature space separately. In the case of the full scenario for the 1M dataset, there is a clear performance-based ordering of the different spaces. As shown in Figure 7.1a, this order is actor, director genre, plot summary emotion, movie review emotion and finally movie space. When the sparsity of the database increases, the best performing spaces (i.e. actor and director) degrades more than the others. This is due to the fact that these spaces have many more features (i.e. the number of actors or directors is many more than the number of genres), and hence are more likely to lack information when sparsity increases. In the extreme case of the cold start problem, director and actor spaces performed the worst.

Also interesting finding is that emotions are useful in sparse scenarios, especially in cold start situations. This is shown by the comparison of movie review emotion with genre spaces (Figure 7.1b). In addition, the comparison between the model using emotion spaces (i.e MPR) and the one using semantic spaces (i.e. MGAD) shows a superiority of the latter one in low sparsity situations (Full in Figure 7.1a). Emotion spaces add an extra dimension of information in high sparsity scenarios as shown with sparsities of 10 and 20 (random and cold start splits) since they decrease MSE and the variance of the performance measure.

Finally, in order to investigate further the importance of the feature spaces, we used the measure of relative influence proposed by Friedman et al. [Fri01] for boosted trees. Each of the features is given an importance between 0 and 100, so that the sum over features equals 100. The importance of a variable reflects how important the drop in performance would be, if the feature was to be removed. We report the box plot of the relative influence over the 10 cross validations for the model (MGADPR) in Figure 7.3.

We can observe different relative influences depending on the amount of data. In the 1M dataset with no sparsity (1M, Full in Figure 7.3), the most important spaces are actor, director, and then genre, and movie. This means that emotion spaces are not important when the data provides sufficient information to predict the rating of the movie based on more direct indicators like the actors in the movie.

However, at the other extreme, i.e. the most sparse scenario (100K, 10 in Figure 7.3), actor, director and movie space do not influence the decision making process of boosted trees as much as the movie review emotion space and genre do. Again we observe that the movie review emotion space acquires more importance than plot summary emotion space.

It can also be observed that there is a connection between the number of different features in a feature space and the use of this feature space for different volumes of data. For example, genre and emotion spaces have less features in comparison to the other ones, and are more likely to be used when there is not much data (i.e. ratings) to train from.

To investigate whether emotions do bring more information than sentiments, we performed experiments using a sentiment feature space[5] instead of emotion feature space (following the method-

[5]The sentiment values were extracted based on the work presented in [SPI08].

ology in Section 7.2.1). The improvement is lower, but not significantly, than that obtained over emotion feature space . However, the usefulness of the emotion feature space is illustrated in the next section where users' latent groups implicitly (through LDA) were created based on the emotions extracted from external movie reviews.

7.4.2 Analysis of Latent Groups

In this section the latent groups discovered by LDA are explained through some examples for two of the most important semantic spaces (i.e. actor and genre) and the emotion spaces. Data learnt by LDA, with the full data (1M, no sparsity) was used for better interpretability.

The most important latent groups based on the expectation of the probability $P(z)$ over the set of training users. For each of these latent groups, we calculated the top and bottom five features f for $p(+|f, s)$ were calculated (see Table 7.1). Note that the number after each feature in the table for movie review (R) and plot summary (P) emotion spaces, corresponds to the discretised quartile to which it belongs (Section 7.2.1), and ranges from 1 to 4. We now discuss our findings, based on the analysis of several latent groups beyond those presented in the table.

In the actor space, we observe that in most of the latent groups, the important actors (i.e. those who play the main role) are separated from other actors (e.g. supporting actors). This is due to the fact that these actors were consistently liked whereas the supporting actors also play in the movies that a user has not liked. The second observation is that the actors who are categorised together in a latent group either play in the same movies or the same genre, or belong to the same period of time. Thus LDA tailed to put together actors into coherent latent groups.

In the genre space, we observed that features within latent groups are also related, e.g., the highest and the lowest probability genres have distinct genre profiles, and the top movies related to this latent group perfectly match the liked genres.

In the movie review emotion space, we observe that the features within the latent group indicate those movies for which users expressed disappointment or dissatisfaction. The reviews for the top movies for this latent group indicate supporting comments such as *"The movie had nothing to do with the title."* and *"It played off more as a B movie."* for *Soft Toilet Seats* and *"The story doesn't* *hit you over the head explaining events like most films"* for *New Rose Hotel*. However, positive comments with high ambiguity such as *"I saw this movie when I was very young and at first never really understood it."* and *"I mistake it for a Disney movie a lot of the time but who can blame me"* can be considered as reproach and therefore the referring movie (in this case *The Swan Princess*) considered in the same group. Therefore, if a user has an unusual taste and likes the movies that the majority of people don't like then more such movies will be recommended to him or her. This feature is unique to this space. On the other hand, if the interest of the user is similar to the crowd, then the recommendation will also be common to that of the crowd such as the movies *Sneakers*, *Amistad*, and *A Simple Plan* for another important latent group.

In the plot summary emotion space, we observe that the features within latent groups meaningfully select movies based on the emotion interpretation of their storyline. For instance, the top latent group corresponds to the movies that have a twist of story or surprising (or shocking) emotion. This makes the plot summary different from genre space. This can be seen better by looking at the top movies for this latent group. For example, *"her character is to be killed off"* or *"Her life begins to fall apart"* for the movie *The Killing of Sister George* or *"A . . . mother . . . facing divorce is thrust back in time . . . Given the chance to change the course of her life . . . "* for the movie *Peggy Sue Got Married*, and *"John Shaft . . . first finds himself up against . . . the leader of the black crime mob . . . finally working with [him] against the white mafia . . . "* for the movie *Shaft*.

In order to gain further insight on the relationship between the different feature spaces, the Pearson's correlation was calculated for the predicted ratings of the model based on single features spaces. Genre and emotion spaces (especially reviews that have a correlation around 0.8 for the complete dataset) had the highest correlation at different levels of sparsity. This is an interesting observation as the information present in these spaces is of a very different nature. However, it can be argued that the emotion expressed in the movie reviews are influenced the most by the genre of the movie. Other spaces had rather low correlations (below 0.4), showing that they are more likely to be complementary. A final observation is that correlation decreases when sparsity increases confirming the observation that considering different feature spaces is important in high sparsity situations.

7.5 Chapter Summary

This chapter investigated the effect of emotion and semantic spaces in improving the performance of a model-based CF system, and analysed the effect of sparsity and dataset sizes in rating prediction accuracy and recommendation precision. Movie and actor spaces were observed to be the most and least sensitive spaces to sparsity and dataset size, and the model that uses all spaces is the best performing one over all sparsities and datasets. It performed better than a state-of-the-art CF system.

The LDA approach to CF was adapted to cluster semantic movie information as well as emotion information based on users' ratings. We proposed to include negative information (movies that have been disliked) into the LDA generation process and for each feature space we calculated the probability that a movie is liked (or not) given a user. This was used to predict a rating using boosted trees.

Results showed that emotional features consistently play a role in improving the recommendation quality in comparison to the scenario where only the movie space (i.e. the baseline) is used. Furthermore, the effectiveness of emotion spaces increase with the sparsity of the dataset, especially in a cold start situation. This indicates that emotion spaces are a potential source of information. A comparison between the improvement achieved in MSE and MAP values shows that emotion spaces are more effective in predicting the actual ratings than detecting top rated movies. We also observed that movie review emotion space and genre space based models predict similar ratings, but it is important to note that emotion features are the outcome of an emotion extraction system and not manually created metadata as is the case for genre, thus they do not require costly and time consuming human intervention.

Emotion extracted from the movie plot summary and movie review emotion spaces affect system performance differently. We believe that this is due to the richer emotional content in opinionated movie reviews than the relatively more objective plot summary texts. It is also important to consider that there is much room for improving the accuracy of emotion extraction techniques.

Table 7.1: The five highest and lowest probability features and the five highest probability of being a feature in a liked movie for the most important latent groups in actor, genre and movie review emotion spaces (No sparsity - 1M dataset)

Space	Features with the highest & lowest probability	Movies with the highest probability
Actor	**Highest:** Tommy Lee Jones, Samuel L. Jackson, Sean Connery, Fred Dalton Thompson, James Earl Jones **Lowest:** Kirk Baltz, Pruitt Taylor Vince, Eddie Izzard, Andy Dick, Steve Lawrence	In the Line of Fire, Die Hard, The Mask of Zorro, U.S. Marshalls, Twister
Genre	**Highest:** Adventure, Crime, Musical, Mystery, Sci-Fi **Lowest:** Comedy, War, Documentary, Action, Western	The Man in the Iron Mask, Brenda Starr, Let's Get Harry, The Avengers, Indiana Jones and the Temple of Doom
Movie Review Emotion	**Highest:** reproach-1, reproach-4, surprise-1, distress-1, reproach-2 **Lowest:** distress-3, joy-2, joy-3, sorry-for-2, resentment-2	Above the Rim, Power 98, Soft Toilet Seats, New Rose Hotel, The Swan Princess
Plot Summary Emotion	**Highest:** surprise-1, shock-3, surprise-2, gloating-4, distress-3 **Lowest:** joy-1, hate-3, joy-3, hate-2, surprise-3	Peggy Sue Got Married, The French Connection, Shaft, Blast from the Past, The Killing of Sister George

Chapter 8

Using Emotion to Diversify Document Rankings

8.1 Introduction

In this chapter we explore the role of emotion features in diversifying document rankings of IR systems. News contain emotionally rich data and several studies have attempted to automatically extract these features from it [MPI07, LYC07]. The results presented in Chapter 6 and 7 showed that using emotion features can improve the effectiveness of a CF system. In this chapter, the effectiveness of emotion features when diversifying document rankings is studied.

We propose to use emotional features to enhance the diversity of the retrieved results since they offer a new way to diversify information, based on emotion, and explore another dimension of emotional relevance. This is motivated by the fact that IR systems strive to gather conceptual information about a document through an indexing process, e.g. by representing documents as a bag of words. However, such a process ignores the fact that documents are not only vehicles for transmitting information, but also convey emotion. Here we propose that diversifying document rankings based on emotion features allows us to better overcome this issue, and in turn improves the overall effectiveness of retrieval systems.

8.2 Literature Survey on Diversifying Document Rankings

Given a query, IR systems rank documents according to their estimated relevance. Searchers' queries are often ambiguous or have multiple facets [SJRS07]. For example *Java* is an ambiguous query since it has different interpretation. Examples of such are the programming language, the island, and the coffee. Differently, *java programming language* is a multi-faceted query since it has several aspects, e.g., development kit download, language specifications, tutorials, courses, and books. Ambiguous or multi-faceted queries are an issue for search engines.

Originally, query ambiguity was not addressed by retrieval algorithms. The independent document relevance assumption, underlying the well-known probability ranking principle in IR [Rob77] might result in retrieving documents for such queries by assuming that they correspond to a single IN. In this case, the searcher's IN can be either fully satisfied (it matches the detected IN) or not at all (it does not). This problem soon became apparent and search engines tried to alleviate it either by predicting the most popular interpretation of the facet of a query using it in their retrieval process (e.g., query reformulation) or explicitly asking users to specify their IN (e.g., query suggestion). However, both query reformulation and suggestion have their shortcomings. With query reformulation there is no guarantee that the searcher's IN is the popular choice, and with query suggestion, there is a risk of not getting any explicit feedback from searchers.

In recent years, diversifying the set of ranked documents has proved to be useful in improving the effectiveness of IR systems [AGHI09] for users. This is because diversity avoids redundancy, resolves ambiguity, increases the coverage of different facts and effectively addresses searchers' information needs [CKC+08]. The systems applying diversification approaches try to maximise the overall relevance of the retrieved results with respect to the facts and/or interpretations of a query while minimising redundancy of considering such aspects [GS09, CKC+08]. The diversification problem is NP-hard [CK06] ; as such, we can only hope to have approximations to this problem; one such approximation is a greedy algorithm, which is employed by most diversification approaches in the literature

In general, diversification approaches can be categorised into explicit and implicit [SPMO10]. Explicit approaches directly model the query aspects in their diversification approach, then maximise the coverage of the selected documents with respect to these aspects. In such approaches, the aspects of a document are gathered from external evidences than from the content of the document itself, e.g., employing a taxonomy for both queries and documents [AGHI09], or using query logs of commercial search engines [RD06, SMO10a] to maximise the coverage of a document ranking with respect to the aspects of a query. This approach shows promising results in improving the effectiveness of IR systems. In contrast to explicit approaches that rely on external evidences to identify the query aspects, in implicit approaches, the similarity between contents of the documents is used for diversification purposes.

Implicit approaches assume that similar documents will cover similar interpretations or facets of a query, and in order to avoid redundancy they should be demoted. In general, they diversify rankings by directly comparing the documents retrieved for a given query to one another, to select those documents that are most similar to the query while being the most dissimilar to the ones already selected. Such approaches are known as the maximal marginal relevance (MMR), the most popular (implicit) diversification approach proposed by of Carbonell and Goldstein [CG98]. Carbonell and Goldstein suggested using any content-based similarity function such as cosine similarity [CG98] to diversify ranking results. Many variations of MMR approach have been proposed that follow a similar principle differing only in how similarity between documents is calculated. For example, similarity can be computed from correlation measures between document [WZ09], Kullback-Leibler divergence measure and a simple mixture model [ZCL03] , or using probabilistic models [CK06].

8.3 Approach

In the following, we explain the diversification approaches used in this work and discuss how emotion features are blended together with estimations of document relevance. Then the construction of emotion vectors is explained.

8.3.1 Diversifying Document Rankings

In order to diversify document rankings, we adopt Maximal Marginal Relevance (MMR) [CG98] as it is an effective and common approach. Let $sim(d, q)$ denote a measure of similarity between document d and query q; this can be regarded as a measure of relevance of d to q. Also let $esim(e(d), e(d'))$ represent the similarity between the emotion vector representations (see Section 8.3.2) of documents d and d'. We consider the situation where $|R|$ documents have been ranked, and the ranking function considers which document has to be ranked next. Following MMR, the next document to be ranked (i.e. d^*) is selected such that:

$$d^* = \arg\max \left[\lambda sim(d, q) - (1 - \lambda) \max_{d' \in R} esim(e(d), e(d'))\right]$$

where λ is a parameter that controls the impact of emotion similarity on the selection of document d^*: e.g. if $\lambda = 1$, emotion similarity has no impact on the selection of documents; while if $\lambda = 0$, emotion similarity is the only criterion used for ranking documents.

We further generalise the MMR approach such that the similarity between the candidate document and the query is interpolated with the *average* emotion similarity between the candidate document and those that have been ranked at previous positions. Thus, under the average interpolation approach (AVG-INT), d^* is ranked at rank position $|R| + 1$ if

$$d^* = \arg\max \left[\lambda sim(d, q) - (1 - \lambda) \sum_{d' \in R} \frac{1}{|R|} esim(e(d), e(d'))\right]$$

In contrast to MMR which considers the most dissimilar among the already selected documents, the AVG-INT approach considers the average similarity between a candidate document and documents ranked in the previous $|R|$ ranks. Several similarity functions can be used for computing $esim(e(d), e(d'))$. In this experiment, cosine similarity and Pearson's correlation are used.

8.3.2 Construction of Emotion Vectors

In order to extract emotions from a retrieved document, OCC1 is used (see Section 5.4) considering the following procedure: Let S denote the set of sentences associated to a document d. The emotion classifier makes a binary decision about each emotion for a sentence, i.e. it decides whether the emotion is present or not. For each sentence we construct a 24 dimension vector using the emotion classifier. Each component can take the value 1 if the emotion is present in the sentence and 0 otherwise. In order to represent the emotion contained in d we average the emotion vectors of the sentences in d to give equal importance to each sentence.

8.4 Experiment and Results

8.4.1 Test Collection

Approaches were tested on the TREC 678 Interactive Track collection[1], which also have been used for diversity task evaluation [ZCL03]. This collection consists of documents belonging to the Financial Times of London corpus, and comes together with 20 topics and associated subtopic relevance judgements. For each of the TREC Topics, we used the title of the topic to generate queries. The collection was indexed using the Lemur toolkit[2]. Standard stop-word removal and stemming techniques were applied at indexing time to both documents and query topics.

8.4.2 Evaluation Procedure

The top n documents (with $n = 20, 50, 100, 200$) were retrieved in answer to each query using a unigram language model with Dirichlet smoothing, where the smoothing parameter was set according to standard values (i.e., $\mu = 2000$) [Zha08]. This run formed the baseline (identified as LM in Table 8.1) against which we compared our diversification approaches. The top n retrieved documents were re-ranked according to our approaches where $esim(e(d), e(d'))$ was computed by the cosine similarity or Pearson's correlation between the emotion vectors representing the documents; while $sim(d, q)$ was estimated according to the scores returned by the baseline. We also tested the MMR approach considering only text features (denoted by *MMR(t)*): this has been implemented according to the MMR equation presented in Section 8.3.1, but using a term vector representation of the documents.

For all the diversification approaches, we varied λ in the range $[0, 1]$ with a step of 0.05. We report the results obtained selecting parameter values that maximise α-NDCG@10 for each query, since

[1] In this preliminary study, we decided to use emotionally rich data to investigate the potential of emotion in diversification task. Therefore we selected the TREC 678 Interactive Track collection (containing news) instead of the TREC Web track collection (containing web pages), since news have been shown to be emotionally rich [LYC07, Aus03] but this cannot be said for web pages.

[2] http://www.lemurproject.org/

We wanted to measure the maximum benefit of using the tested approaches. Results were similar both when using the cosine similarity and the Pearson's correlation.

8.4.3 Metrics

Two official metrics are used to evaluate the ranking approaches: the α-normalised discounted cumulative gain α-nDCG [CKC+08] and the intent-aware precision (IA-P [AGHI09]). The parameter α in α-nDCG balances the importance of relevance and diversity. Tuning the value of α changes the rewarding strategy of the metrics for diversity and relevance. When the value of α is high (e.g., close to one), diversity is rewarded more, and when this value is low (close to zero) the relevance is. In the case where $\alpha = 0$ this metric is equivalent to the traditional nDCG [JK02].

The IA-P metric accounts for the possible aspects underlying a query and their relative importance. This metric extends the traditional notion of precision to reward the diversity of a ranking.

Both α-nDCG and IA-P are computed following common practice [CCS09]: α-nDCG is computed with $\alpha = 0.5$, in order to give equal weights to both relevance and diversity and IA-P is computed with all query aspects considered equally important. In addition, both α-nDCG and IA-P are reported at three different rank cutoffs: 5, 10, and 20. These cutoffs focus on the evaluation at early ranks, which are important in the task of diversity [JSBS98].

8.5 Results and Discussion

8.5.1 Quantitative Study of the Performance of the Models

Table 8.1 reports α-nDCG and IA-precision values of Language Model (LM), MMR with text features (MMR(t)), MMR with emotion features (MMR(e)) and AVG-INT with emotion features (AVG-INT(e)), and presets percentages of improvement over LM (i.e., baseline) in parenthesis. The best performing approach at each rank is highlighted in bold. For MMR(t) we reported the best performing results (i.e. top 100 documents). Note that other settings obtain results that exhibit similar trends. Performance of AVG-INT with text features and MMR(t) are similar, and we

Table 8.1: α-nDCG values of Language Model (LM), MMR with text features (MMR(t)), MMR with emotion features (MMR(e)) and AVG-INT with emotion features (AVG-INT(e)) are reported and percentages of improvement over LM are presented in brackets. The best performing approach at each rank is highlighted in bold.

		LM	***MMR(t)***	***MMR(e)***			***AVG-INT(e)***		
			Top 100	Top 20	Top 50	Top 100	Top 20	Top 50	Top 100
α-nDCG	@5	0.520	0.549 (+5%)	**0.568 (+9%)**	0.555 (+7%)	0.545 (+5%)	0.561 (+8%)	0.559 (+8%)	0.539 (+4%)
	@10	0.532	0.565 (+6%)	0.560 (+5%)	**0.567 (+6%)**	0.551 (+4%)	0.554 (+4%)	0.555 (+4%)	0.547 (+3%)
	@20	0.545	0.569 (+4%)	0.556 (+2%)	0.564 (+4%)	0.546 (+0%)	0.555 (+2%)	**0.565 (+4%)**	0.559 (+3%)
IA-p	@5	0.103	0.105 (+3%)	**0.111 (+8%)**	0.108 (+5%)	0.106 (+4%)	0.108 (+5%)	0.105 (+3%)	0.104 (+1%)
	@10	0.077	0.081 (+5%)	0.083 (+8%)	**0.083 (+8%)**	0.081 (+5%)	0.082 (+6%)	0.081 (+5%)	0.080 (+4%)
	@20	0.053	0.056 (+5%)	0.053 (+0%)	0.056 (+5%)	0.055 (+4%)	0.053 (+0%)	**0.057 (+7%)**	0.055 (+4%)

therefore report the latter. Because the difference in performance and its variance is large (small topic set composed of 20 queries), performing significance tests would not be appropriate [vR79, pages 178–180].

The results show that considering emotion features improves retrieval effectiveness since emotion-based approaches display better performances than LM (i.e., baseline). We found that emotion-based diversification obtained substantial gains in terms of α-nDCG (about 20%) for more than 30% of queries over LM. They also provide better performance than the *MMR(t)* approach, which only employs text features. The results observed from α-nDCG and IA-precision show similar patterns of improvement in the effectiveness of the system.

In general, diversifying ranking based on emotion features are more effective when smaller number of documents are taken into account (e.g., top 20 performs better than top 50 and 100). The reason lies in the fact that the lower the position of a document in the original rank list (e.g., document at position 100 compare to the one at position one), the higher the chance that the document is not relevant, thus, the more chance of boosting its position through diversification process based on emotion features. Since the performance of the system considering top 20 documents is the best one for the rest of the analysis we focus on this scenario.

Comparison between MMR and AVG-INT approaches on emotion features at top 20 show that MMR performs better than AVG-INT at all the evaluation cut-offs (i.e., 5, 10, and 20). The difference between MMR and AVG-INT approaches is in their method for selecting the next relevant candidate document. MMR uses the maximum value of dissimilarity between the candidate document and the documents in the list whereas AVG-INT uses the average of dissimilarities between the candidate documents and the documents in the list. The better performance of MMR over AVG-INT indicates that for emotion features being very dissimilar to one document works better than being dissimilar to overall documents in the list.

This section presented a qualitative analysis of the effectiveness of considering emotion features to diversifying document rankings. Whilst the average effectiveness gains are marginal in this preliminary study, there is a case for using emotion features to diversifying document rankings. We also observed that the effectiveness of the emotion-based diversification approaches is query dependent, therefore, there is a case for selective diversification [SMO10b].

8.6 Chapter Summary

In this chapter, we investigated the effectiveness of using emotion features when diversifying document rankings. Two research questions were studied in this experiment: (1) is the emotion representation (emotion object) of news a good representation for diversifying the retrieved results, and (2) does this representation improves performance. The results are encouraging and show improvement when including emotion features for re-ranking retrieved results, compared to using pure textual features.

In order to represents news, emotions were extracted from news articles themselves and used as their emotion representation. The results show that some topics gain more from emotion-based diversification than others. It also show that the retrieval effectiveness for queries varies when the diversification approach is based on textual representation of the retrieved results or based on emotion representation. This shows that an emotion representation of news gives a novel view over news articles, complementary to the standard text representation used in IR. This work is a foundation for future research on utilising emotion features in IR tasks.

Part IV

CONCLUSIONS

Chapter 9

Conclusions and Future Works

9.1 Contributions and Conclusions

In this section, we list the contributions of this thesis, and outline the main conclusions.

9.1.1 Contributions

This thesis investigated the benefits of considering emotion for the purpose of Information Retrieval (IR). It argued that by considering emotion: (a) IR can be founded on a more realistic understanding of the searcher and search processes; (b) IR systems built upon this can better address searcher's needs; and (c) this in turn can lead to more effective search systems.

We first present the overall contribution before going into more detail.

1. By reviewing literature from IR/IS and also philosophy, we have proposed to incorporate emotion as both a primary and secondary factor in the IR&S process.
2. We have revised the fundamental concepts of IR, such as searchers' need, document representation, and relevance, in Chapter 4, taking into account the shortcomings in the light of contributions above, and have introduced three new concepts, namely emotion need [a desire to be in a particular emotion state], emotion object [an emotion representation of

a document] and emotion relevance [a relation between emotion aspect of documents and emotion need].

3. We have presented a conceptual map of IR scenarios in Chapter 4, in which these three concepts play important roles, and have described the main pragmatic challenges facing emotion-based IR systems.

4. We have developed an open-source text-based emotion extraction system and analysed its performance exhaustively.

5. We have developed two novel collaborative filtering models that incorporate emotion as well as semantic features of movies and their ratings. Experiments showed that considering emotion features not only improves the effectiveness of the system but also alleviates the data sparsity problem, especially in the cold start item problem.

6. We have developed a news retrieval system which diversifies results based on their emotion aspect and have showed that the emotion-based diversification approach improves the retrieval effectiveness.

In the following, we summarise the main concepts discussed in the thesis, along with the main theoretical and practical contributions. In particular, we focus on results that relate to the conceptual map presented in Chapter 4. This chapter concludes with a summary discussion of the implications of this thesis.

9.1.2 Conclusion

The main conclusion of this thesis is that *emotion plays a key role in IR and it is important to consider this*.

- The collaborative filtering (CF) system performs better when it takes emotion features associated with the items into account than when only rating information is considered (Chapter 6 and 7).

- The data sparsity problem associated with CF systems can better be alleviated when the emotion features are taken into account than when only rating information is used (Chapter 7).
- The cold start problem associated with CF systems can better be addressed when the emotion features are taken into account than when semantic features or rating information are (Chapter 7).
- Considering emotion information in addition to semantic and rating information further improves the accuracy of the system compared to the scenario where rating and semantic information were used (Chapter 7).
- The retrieved results better cover the sub-topics associated with a query when they are diversified based on their emotion feature than the scenario when there is no diversification applied (Chapter 8).
- For some topics, the diversification based on emotion performs better than the scenario where it is diversified based on their text representation (Chapter 8).

In the following, I elaborate the theoretical and practical contributions further.

9.1.3 Theoretical Contributions

This section summarises the main theoretical contributions leading to the conclusion of this thesis. We explain these contributions with respect to three main components of IR&S process, namely the searcher's need, document representation, and IR models.

Contribution 1: Emotion act as both primary and secondary factor in IR&S behaviour

The primary factor of emotion refers to the fact that emotion can be considered as an individual need which can motivate searchers to engage in an IR&S process. The secondary factor of emotion refers to the fact that emotion (in relation to cognition) influences every aspect of the searchers' IR&S behaviour, and can thus influence the success or failure of an IR&S process. First, we will elaborate on emotion as a secondary factor in IR&S process.

As discussed in Section 2.4, Kuhlthau [Kuh93] and Nahl [Nah05] investigated the secondary nature of emotion in IR&S scenarios. They observed that participants experience a burst of negative feelings due to uncertainty associated with vague thoughts, leading them to recognise that they have an information need. They agree that there is a positive correlation between a successful information seeking process and a decrease in these negative feelings [Kuh93]. From this point of view, emotion is a factor that exists throughout an IR&S process which aims to meet an IN.

In this thesis, emotion as a secondary factor in IR&S process is supported further by presenting research from other domains such as philosophy and psychology. Research in philosophy showed that emotions arise as a result of an evaluative judgement (see Section 3.2.2). We argued that the IR&S process, being a cognitive process, includes many evaluative judgements, and that emotions therefore arise throughout the IR&S process (see Section 3.2.3). Research in psychology showed that emotion affects every aspect of human behaviour, including rationality, decision-making, attention, and memory-access. Several studies reported that emotion plays an important role in such behaviour [Sou90, Dam94, GAT98, SS88]. For example, without emotion, decision-making [Sou90] and attention [LeD96] are not possible (see Section 3.4). Considering emotion as a secondary factor in an IR&S process was therefore supported by such theories.

However, we argued in this thesis that emotion acts as a primary factor in an IR&S process, i.e., we consider emotion as autonomous and important. This argument was supported mainly by research in sociology. This is one of our theoretical contributions to IR/IS since the primacy of emotion has been mainly ignored in these domains.

The first supporting evidence for our argument is that human beings seek each other for informational and emotional support when they are in emotionally-challenging situations (i.e., emotion need) (see Section 3.5.3). Such behaviour is called coping [LL94]. Researchers show that, in the digital age, most social behaviour, including coping, manifests itself on the internet [LEG01]. IR applications are one of the most important applications on the web, and have the purpose of satisfying needs, and are thus central to support such behaviour. It is natural to consider coping as one of the main process in the IR&S behaviour.

The second supporting evidence is based on mood-management theories (see Section 3.5.4). In Section 3.5.4, we discussed hedonistic motivations underlying entertainment activities, which are particularly present in media consumption. Evidences for the use of IR applications for such behaviour is apparent and ever-increasing, e.g., the popularity of movies, music, news, etc. People may observe a set of documents/contents which may emotionally stimulate them in a way that satisfies their emotion need, and can thus be considered as relevant [Oli03]. For this reason, considering emotion as a primary factor in an IR&S process is justified.

We explained that in the normative view of IR/IS, as discussed in Section 2.4, the focus is on the satisfaction of searchers' IN. Although the role of emotion is acknowledged as a factor influencing the whole IR&S behaviour, its role was limited to the study of its influence on the process of satisfying an IN. However, as explained in Section 3.5, emotion can be a source of motivation on its own for a searcher to engage in an IR&S process. Such scenarios have not been considered in the IR/IS community, and this motivated the reframing of searcher's system of needs (see Section 4.2) and the definition of the *emotion need* concept.

Contribution 2: Emotion Need is the central need in the searcher's need system

The centrality of emotion need in the searcher's need system is supported by literature from both IR/IS and sociology. In IR/IS, as discussed in Section 2.4, researchers explain the possibility that an emotion initiates an IN. This has been discussed from two perspectives:

1. works that tend not to explicitly refer to an emotion need and discuss the role of human emotion in an information seeking process, e.g., Kuhlthau's model [Kuh93] suggests that the need of the searcher is to be satisfied and that this happens when the negative feelings such as anxiety, distress, etc. associated with an uncertainty diminish through IR&S processes.

2. works that consider emotion as an individual need, e.g., Wilson [Wil93], who suggests that this need can either directly motivate a searcher to engage in an IR&S behaviour or initiate an information need processes.

Based on the above, we proposed that an emotion need underlies any information need (i.e., the need for satisfaction or more specifically the need to diminish negative feelings). This is also supported by sociology research such as coping theory (see Section 3.5.3). Nahl [Nah05] explains that IR&S processes can be categorised as a coping process. Coping theory explains that people employ different strategies, e.g., seeking information, to diminish their negative feelings such as stress, anxiety, etc. We proposed that IR&S processes, used when searching to satisfy an IN are, more precisely, part of the *problem-focused strategies"* (see Section 3.5.3).

However, we argued that there are emotion needs that can motivate searchers to engage in IR&S behaviour which strictly speaking does not have an IN (see Section 3.5). This argument is supported by other types of coping strategies, such as "appraisal-focused" and "emotion-focused" (see Section 3.5.3). In these strategies people seek for evaluative judgements (e.g., opinion) and/or emotion to diminish their negative feelings [Laz91]. People employ coping strategies such as arousal-focused and emotion-focused (Section 3.5.3) in their daily lifestyle [LEG01]. This is due to the pervasiveness of social applications on the web (such as Facebook and Twitter) and the diverse range of documents (movies, music, images, news, blogs, etc) which make such infrastructure available on the web to satisfy emotion need.

On the other hand, other researchers explain the hedonistic motivation behind the consumption of entertainment on the web (see Section 3.5.4). For example, mood-management theories [Zil88b, Zil88a] suggest that searchers, consciously or subconsciously, work to satisfy other emotion needs, i.e., maintaining or changing their emotional state with a hedonistic motivation [Oli03]. In addition to hedonistic motivation, Knobloch [Kno03] shows that there are other emotion motivations which do not have hedonistic motivation. She shows that people sometimes try to maintain particular emotions (usually negative) for a long-term benefit, e.g., keeping the anger for an upcoming conflict.

The evidence mentioned above forces us to rethink and acknowledge the existence of emotion need as a central need in IR/IS. We argued that documents can carry both information and emotion, and that the emotion part of a document can satisfy an emotion need, especially in the case where there is no specific information need (Section 4.3). This was supported by research in sociology. For example, Zillmann [Zil88a] shows that the emotion aspect of documents is an important hedonistic motivation for searchers. Knobloch [Kno03] shows that the emotion aspect of the news is an important criterion for participants to maintain their negative emotions.

The fact that people can communicate emotion via documents is explained by empathy (see Section 3.5.2). Empathy [Dav06] explains that human beings are capable of understanding and experiencing the thoughts and feelings of another person. Due to the same capability, one also may experience emotion while watching an imaginary story presented in movies, or observing contents created by another person such as an image, a text, etc. [Oli03] Therefore, empathy allows people to communicate emotion to each other either directly (e.g. face-to-face interaction) and/or indirectly (e.g. through text).

Contribution 3: Emotion extracted from the document represents document itself

In Section 4.3, we proposed and discussed the fact that the emotional content of documents can be faithfully and usefully extracted from their text. We argued that the extracted emotions represent the document itself (see Section 4.3.3). We justified our argument by contrasting with two other possible cases, i.e., extracted emotions: (1) represent the emotion of the creator of the document while making it; or (2) represent the emotion of the observer of the document when interacting with it.

We argued that the emotion extracted form a document does not necessarily represent the emotion of its creator because extracted emotion represents the message that authors of documents transmit rather than the authors' emotion while writing them (see Section 4.3.1). We described two reasons underlying the *possible difference*[1] between the emotion transmitted/conveyed in a document and the emotion experienced by the creator while making the document: (1) people often may not express their true feelings [Sul02], therefore, they express emotion differently to what they really experience; and/or (2) the emotion experienced by the creator while making the documents sometimes is not important or relevant to the message that he or she wants to transmit. For example, a journalist who is under pressure to finish an article (i.e., experiencing stress), does not necessarily experience an emotion similar to the emotion of the message about an improvement in a new health care system which he tries to report in his article. In this case, his emotion can be considered irrelevant to the message.

We also argued that the emotion extracted from a document cannot represent the observers' emotion, because the assumption that all observers experience emotion similar to the extracted emotion

[1]We do not argue that the emotion extracted from the document never matches the emotion of its creator while making the document, but we argue that there could be situations where these two emotions do not match.

from the document is obviously wrong (see Section 4.3.2). Emotion features captured, for example, via monitoring readers' biometric features and/or facial expressions show the subjectivity of the emotional experience [AAJ10]. Observers can not only experience emotion more or less independently from those extracted from documents but also experience different emotions from one another.

We argued that the emotion experienced by observers is neither objective nor completely subjective, but is intra-subjective (see Section 4.3.2). If emotion were completely subjective, then it would be impossible, for example, for the writers to write something since they would not anticipate what message the readers would get after reading their document. What makes such communication possible is the knowledge of the creator of a document about his/her audiences. For a general audience, it is argued that the creators' knowledge is common-sense (cultural knowledge). To this category belong the majority of objects: movies, music, and many textual documents whose audiences are general public, e.g., books, news, etc.

It has been explained in Section 4.3.3 that the emotion aspect of documents is important in satisfying searchers' needs, and particularly to their emotion needs. This is supported by mood-management theories [Zil88a], which explain that one of the important factors in satisfying hedonistic motivation is the emotional aspect of the documents (see in Section 3.5.4) . Also, in emotion-focused and arousal-focused coping strategies [Laz91] (Section 3.5.3), the emotion aspect of the documents is explained as being important. In such scenarios, a desired emotional experience is a searcher's crucial factor in deciding if a document is relevant or not [Oli03]. Therefore, the emotional aspect of documents plays an important role in IR&S scenarios. We therefore expanded the traditional concept of relevance, and introduced the notion of emotion relevance. Emotion relevance defines how the emotion aspect of a document is important to satisfy a searcher's emotion need.

Contribution 4: The emotion aspect of documents is an important feature in satisfying searchers' needs

In Section 4.4, we described the importance of the emotion aspect of documents from two perspectives, that of the searcher and the system. From a searcher perspective, it has been argued that the emotion aspect of documents can be relevant/interesting to both searchers' information and emotion need. We argue that searchers' IN can be explicitly or implicitly about the emotion aspect of documents (Section 4.4.1). In explicit scenarios, the emotion inclination of the information need is explicitly expressed (e.g., a search for negative emotions in a new product review) whereas in implicit scenarios, the searchers are either unaware of this emotional affinity, or they do not explicitly express it (e.g., the history of a searcher's interactions indicates documents as relevant that express disadvantages of the new health care system without his query explicitly expressing this criteria). In such scenarios, the emotions of the document can either directly or indirectly satisfy a searcher's information need. Emotion need can also be explicitly or implicitly about the emotion aspect of documents (Section 4.4.2).

From a systems point of view, we explained the role of emotion relevance in responding to searchers' needs. We proposed two sources of evidence on which an IR system can rely in order to decide the appropriateness of emotion relevance (see Section 4.4.3).

First, it is proposed that the information captured from searchers' interaction with IR systems, either as explicit or implicit feedback, stored as a profile, is an important source of evidence for such systems to make their decision upon. If a searcher's profile suggests a correlation between what a searcher considers interesting or relevant, and the emotion aspect of the visited documents, then the system, which incorporates emotion relevance in its relevance judgement algorithm, can more accurately and effectively retrieve or recommend documents for the searcher. The correlation could be estimated from documents such as reviews associated with the visited information objects (e.g., movies watched). The proposed idea was extensively addressed in the context of collaborative recommendation in Chapters 6 and 7 and it was shown that emotion relevance improves the accuracy of rating prediction in such systems.

Second, the characteristics and nature of the data (or datasets) could suggest a particular emotion affinity. For example, it would be possible that documents considered to be topically relevant to a particular topic be diversified with respect to their emotion aspects, and thus cover a better range of (emotion) subtopics. This was explored in Chapter 8, in which preliminary experiments have shown that emotion aspects of documents can indeed have such potentiality.

9.1.4 Practical Contributions

This section summarises the main practical contributions leading to the conclusion of this thesis. The four main practical contributions of the thesis are as follows: a text-based emotion extraction system, two movie recommender systems that integrate emotion information along with the semantic and rating information in their rating prediction, and a news retrieval system that diversifies results based on their emotion aspect.

Contribution 5: Text-based Emotion Extraction System

We implemented an open source text-based emotion extraction system. Extracting emotion from text was the most important pragmatic challenge associated with the practical study of the role of emotion in IR scenarios (see Section 4.5). Although emotion-based interaction is more pervasive for media content, there are several important applications involving texts such as news media, advertising, and text that accompanies audio/video content. A proper integration of emotion into IR/IS systems thereby starts with an investigation of the emotion content of text. In this thesis, we focused on textual documents.

In Chapter 5, we investigated the accuracy of the implemented system compared with two other commercial emotion extraction systems. The results of the experiments showed that our implementation of emotion extraction method originally proposed in [Sha08], called OCC1, was more accurate in terms of precision and F-measure than other text-based emotion extraction techniques. Therefore, OCC1 was used to extract emotion in the experiments conducted during this thesis.

We also reported extensive experiments on the effect of parameters on OCC1 performance. The effectiveness of the system was reported with the F-measure, for each possible setting of the parameters. The main and interaction effects of these features were investigated. The results obtained from the main affect analysis indicated that the base list and triplet extractor have great impact on the performance. The results obtained from the interaction effect analysis showed that the OCC1 which employs all features results in the best-performing system.

Contributions 6 & 7: Movie Recommender Systems

Two movie recommender systems were built to investigate a scenario where *searchers have needs (emotional or informational) which are not explicitly identified and the emotion objects can be indirectly useful* as a branch of the conceptual map proposed in Section 4.4.

Movies are one of the most popular sources of entertainment and they are emotionally rich. People watch movies to satisfy their emotion need. There are many theories that try to explain the strong emotional effect of movies on their audiences, such as mood-management and disposition theories (see Section 3.5.4). Audiences of a movie are usually exposed to a longer period of emotional content compared with other multimedia contents (e.g., one or two hours for movies in comparison to a few minutes in video clips). This longer exposure results in a stronger emotional experience in the audiences. The audiences can go through all sorts of emotional experiences; they can get emotionally attached to the story and the main character, and experience a great gratification or sorrow if something good or bad happens to the main character at the end of the movie (Section 3.5.4).

All of these emotional experiences and their duration make the audiences quite sure about their experience of the movie. Therefore, the reviews they provide, based on their emotional experiences, are a valuable source of information, and can be used to represent the movie.

To exploit this, we considered the emotion extracted from the movie reviews as a source of information. In addition, movies are developed for general audiences (see Section 4.3.2); thus, their story is expressed based on a common-sense background. Their plot summaries should thus reflect the emotion of the movie. Therefore, we also considered the extracted emotion from plot summaries to represent the movies.

These two emotional representations (i.e., emotion objects) of the movies were used to improve the effectiveness of CF systems. CF systems suffer from data sparsity since there is not enough information available to such systems for reliable rating prediction (Section 6.3). The use of additional information has shown improvements in the accuracy of CF systems [MAPJ09]. Therefore, such systems are good candidates for the investigation of the effectiveness of emotion as an additional source of information.

In Chapters 6 and 7, we conducted two sets of experiments with two different models with the overall aim of studying two research questions: (1) whether emotional information is useful to improve the effectiveness of the rating prediction in CF tasks in comparison to the situations where only the rating data is used; and (2) whether this information brings additional (complementary) information to what exists so far for movies (actors, director and genre).

In first model, for each semantic and emotion space we propose to construct two vector spaces and define the representation of users and items, so to be able to compute distances in those spaces. Finally, in order to predict a rating, the information about the different spaces is aggregated using the extension of Wang et al. [WdVR08] unified relevance framework. (Chapter 6)

In second model, for each semantic and emotion space we propose to construct latent groups of users. In order to do so, we extend a well-known model-based approach, namely Latent Dirichlet Allocation (LDA) [BNJ03] to avoid its popularity problem. In each space, we propose a methodology to compute the probability that a given user likes an item. Finally, in order to predict a rating, the information about the different spaces is aggregated using standard machine learning techniques. (Chapter 7)

The results of both experiments showed that the emotion extracted from movie reviews and plot summaries improved the accuracy of CF prediction. Emotion extracted from movie reviews performed better than plot summaries. We believe that this is due to the richer emotional content in opinionated movie reviews than the relatively more objective plot summary texts. The results of the experiment presented in Chapter 7 showed that emotion information further improves the performance of CF systems when they are combined with the scenarios where semantic information such as actor, genre, and director has already been used.

However, the improvement achieved by the emotion representation of the movies depends on the quality of the employed emotion extraction system. The study of its effectiveness presented in Chapter 5 showed that, although the extraction system performed better than other available emotion extraction systems, there is still room for its improvement. Therefore, we discuss the finding that, by improving the accuracy of emotion extraction systems, it is possible to improve the emotion representation of the movies, and, in turn, to improve the performance of the CF systems.

In addition, a more elaborate method for using the emotion of the movie reviews and plot summaries can further improve the accuracy of the CF systems. So far, we aggregated the emotions in the movie reviews and plot summaries regardless of their intentionality, but using a more sophisticated and better way of representing emotion objects can also improve the performance of the CF systems. One approach could be to consider the object that the emotion refers to in the text [JWMG09]. For example, in the case of movie reviews, for the sentence *"I was unhappy with the ending of the movie"* the unhappy emotion refers to the end of the movie, whereas for the sentence *"I was unhappy with the way this actor played the role"* the unhappy emotion refers to the portrayal given by a particular actor. By considering the object that emotion refers to, we can potentially better represent movies.

Contribution 8: News Retrieval System

The news retrieval system was built to investigate a scenario in which *searchers have information needs that are explicitly expressed via a query but the emotion object can be used implicitly and indirectly to help users to better satisfy their need.* This corresponds to another branch of the conceptual map for emotion relevance (Section 4.4).

Compared to the movie recommender system, in the news retrieval system there is no prior information available and/or captured from the searcher interaction (i.e., no profile) at the time of retrieval. Therefore, in this experiment, system's knowledge of the searchers' IN is limited to the searchers' submitted query.

However, as explained in Section 8.2, queries are usually ambiguous or multifaceted and a promising approach to solve this problem is to diversify the retrieved results so that they cover all possible aspects of a query. The diversification of the retrieved results increases the effectiveness of the system by increasing the chance of addressing the searcher's need.

News helps people to understand their society and surroundings, as well as events around the globe. Despite of the general belief that news contains only factual information, many researchers have explained the importance of their emotional aspect in satisfying emotion need (consciously or subconsciously). For example, researchers have shown that people use news for coping [See04], mood-management [Aus03], and maintenance of strategic emotion (e.g., anger) [Kno03]. Therefore, the emotional aspect of news is an important factor for disambiguating searchers' queries in a news retrieval scenario.

The difference from the CF task is that in this task the emotion representation of news is based on the emotion extracted directly from news (i.e., document level) whereas in the movie scenario the emotion representation of movies is based on the emotion extracted from movie reviews and plot summaries (i.e., collection level). That is, the emotion representation of news is extracted from the same text that is used to represent news for its textual representation, whereas in the movie scenario the emotion representations and the semantic representations of movies came from different sources of information.

In summary, the research questions were: (1) is the emotion representation (emotion object) of news a good representation for diversifying the retrieved results?; and (2) does this representation improve performance?

For that purpose, in a preliminary study, we adapted a well-known diversification approach to using emotion representation rather than textual one (as originally proposed). The performance of the system was then compared to situations where no diversification and a text-based diversification were applied. We propose to use emotional features to enhance the diversity of the retrieved results since they offer a new way to diversify information, based on emotion, and explore another dimension of emotional relevance. (Chapter 8)

The results show that the diversification of the retrieved results improved the effectiveness of the system. They also show that some topics gain more from emotion-based diversification than others. More importantly, the effectiveness of the diversification approach varied between textual representation and emotion representation, showing the fact that emotion representation of news articles captures different aspect of them than their textual representation does.

Although the overall improvements were marginal, the results were encouraging. Possible ideas for further improving the system are as follows: similarly to what we have previously mentioned, improving the accuracy of the emotion extraction system may result in better effectiveness of such systems. It is also important to design more elaborate models that use both emotion and textual representation of data in their diversification approach.

9.2 The Way Ahead

In this thesis, we have opened a new line of research in IR. This being a very fundamental approach, the new directions for research suggested by this thesis are wide. A complete scheme of the role of emotion in IR&S behaviour was put together and explored, but there are still a number of theoretical and experimental aspects that are to be studied further.

9.2.1 Directions for Theoretical Research

In this thesis, following the traditional view of IR, an assumption has been made: it is not possible for IN, from an IR system point of view, to know what the emotion need of the user is. What an IR system can understand is the evidence gathered from interaction, such as submitted queries, implicit or explicit feedback, and searchers' profiles. A future direction of research could be to investigate whether it is possible to understand the motivation of the searcher to engage into an IR&S process, e.g., whether it is because of an information or emotion need. If such recognition is possible, the next step could be to develop a richer understanding of their motivation, e.g., whether it is coping or mood-management.

9.2.2 Directions for Practical Research

This thesis investigated the practicality of only a small part of the proposed conceptual map of emotion relevance presented in Section 4.4. This is due to two main reasons: (1) investigating all possible scenarios explained in the conceptual map is a long-term task; (2) some scenarios were more difficult to investigate than others due to limitations associated with them, such as lack of dataset, evaluation methodology, metrics and procedure. An attempt to solve such limitations is a possible direction for research, as well as the investigation of the practicality of scenarios explained in the conceptual map.

9.3 Chapter Summary

This thesis concludes, that *emotion plays a key role in IR and it is important for it to be considered*. The main theoretical and practical contributions of the thesis that led to the thesis conclusion were presented. At the end, possible avenues for future work were proposed.

Bibliography

[AAJ10] Ioannis Arapakis, Konstantinos Athanasakos, and Joemon M. Jose. A comparison of general vs personalised affective models for the prediction of topical relevance. In *Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 371–378, 2010.

[AGHI09] Rakesh Agrawal, Sreenivas Gollapudi, Alan Halverson, and Samuel Ieong. Diversifying search results. In *Proceedings of the 2nd ACM International Conference on Web Search and Data Mining*, pages 5–14, 2009.

[AJG08] Ioannis Arapakis, Joemon M. Jose, and Philip D. Gray. Affective feedback: an investigation into the role of emotions in the information seeking process. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 395–402, 2008.

[AKJ09] Ioannis Arapakis, Ioannis Konstas, and Joemon M. Jose. Using facial expressions and peripheral physiological signals as implicit indicators of topical relevance. In *Proceedings of the 17th ACM International Conference on Multimedia*, pages 461–470, 2009.

[AMJ+09] Ioannis Arapakis, Yashar Moshfeghi, Hideo Joho, Reede Ren, David Hanh, and Joemon Jose. Relevance estimation by facial expression analysis. In *Proceedings of IEEE International Conference on Multimedia and Expo*, pages 29–37, 2009.

[Arn60] Magda B. Arnold. *Emotion and Personality*. Columbia University Press, 1960.

[AS05] Cecilia Ovesdotter Alm and Richard Sproat. Emotional sequencing and development in fairy tales. In *Proceedings of the First International Conference on Affective Computing and Intelligent Interaction*, pages 668–674, 2005.

[AT01a] Gediminas Adomavicius and Alexander Tuzhilin. Expert-Driven Validation of Rule-Based User Models in Personalization Applications. *Data Mining and Knowledge Discovery*, 5(1):33–58, 2001.

[AT01b] Gediminas Adomavicius and Alexander Tuzhilin. Multidimensional Recommender Systems: A Data Warehousing Approach. *Electronic Commerce*, 2232:180–192, 2001.

[AT05] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.

[Aus03] Charles F. Aust. Factors in the appeal of news. *In Jennings Bryant, David Roskos-Ewoldsen, & Joanne Cantor (Eds.), Communication and emotion: Essays in honor of Dolf Zillmann*, pages 511–530, 2003.

[Bat91] Charles Daniel Batson. *The altruism question: Toward a social psychological answer*. Lawrence Erlbaum, 1991.

[Bed56] Errol Bedford. Emotions. In *Proceedings of the Aristotelian Society*, volume 57, pages 281–304, 1956.

[BES10] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the 7th Conference on International Language Resources and Evaluation*, pages 2200–2204, 2010.

[BM84] Andrew G. Billings and Rudolf H. Moos. Coping, stress, and social resources among adults with unipolar depression. *Journal of Personality and Social Psychology*, 46(4):877–891, 1984.

[BNJ03] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, 2003.

[BNW74] Ehor O. Boyanowsky, Darren Newtson, and Elaine Walster. Film preferences following a murder. *Communication Research*, 1(1):32–43, 1974.

[Bow81] Gordon H. Bower. Mood and memory. *American psychologist*, 36(2):129–148, 1981.

[BS97] Marko Balabanović and Yoav Shoham. Fab: content-based, collaborative recommendation. *Communications of the ACM*, 40(3):66–72, 1997.

[BYRN99] Ricardo A. Baeza-Yates and Berthier Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., 1999.

[Can27] Walter B. Cannon. *Bodily Changes in Pain, Hunger, Fear and Rage*. Cannon Press, 1927.

[CCS09] Charles L. A. Clarke, Nick Craswell, and Ian Soboroff. Preliminary report on the Trec 2009 Web Track. In *Proceedings of Text Retrieval Conference*, 2009.

[cDMMM06] Marie catherine De Marneffe, Bill Maccartney, and Christopher D. Manning. Generating typed dependency parses from phrase structure parses. *Linguistics in the Netherlands*, 20(14):449–454, 2006.

[CdRN09] Valeria Carofiglio, Fiorella de Rosis, and Nicole Novielli. Cognitive emotion modeling in natural language communication. *Affective Information Processing*, pages 23–44, 2009.

[CG98] Jaime Carbonell and Jade Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 335–336, 1998.

[CK06] Harr Chen and David R. Karger. Less is more: probabilistic models for retrieving fewer relevant documents. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 429–436, 2006.

[CKC+08] Charles L.A. Clarke, Maheedhar Kolla, Gordon V. Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. Novelty and diversity in information retrieval evaluation. In *Proceedings of the 31st annual international ACM*

SIGIR conference on Research and development in information retrieval, pages 659–666, 2008.

[CO00] Gerald L. Clore and Andrew Ortony. Cognition in emotion: Always, sometimes, or never. *Cognitive Neuroscience of Emotion*, pages 24–61, 2000.

[COF87] Gerald L. Clore, Andrew Ortony, and Mark A. Foss. The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology*, 53(4):751–766, 1987.

[CSC94] Gerald L Clore, Norbert Schwarz, and Michael Conway. Affective causes and consequences of social information processing. *Handbook of Social Cognition: Basic Processes*, 1:323–417, 1994.

[CSW89] Charles S. Carver, Michael F. Scheier, and Jagdish Kumari Weintraub. Assessing coping strategies: A theoretically based approach. *Journal of Personality and Social Psychology*, 56(2):267–283, 1989.

[Dam94] Antonio R. Damásio. *Descartes' error: emotion, reason, and the human brain*. Avon Books, 1994.

[Dar56] Charles Darwin. The expression of the emotions in man and animals. *The American Journal of the Medical Sciences*, 232(4):477, 1956.

[Dav83] Mark H. Davis. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1):113–126, 1983.

[Dav94] Mark H. Davis. *Empathy: A social psychological approach.* Westview Press, 1994.

[Dav06] Mark H. Davis. Empathy. *In Jan E. Stets & Jonathan H. TurnerHandbook (Eds.), the Sociology of Emotions*, pages 443–466, 2006.

[Dil98] James Price Dillard. The role of affect in communication, biology, and social relationships. *In Peter A. Andersen & Laura K. Guerrero (Eds), Handbook of communication and emotion: Research, theory, applications, and contexts*, 1998.

[Duf62] Elizabeth Duffy. *Activation and Behaviour*. Wiley, 1962.

[ED94] Paul Ekman and Richard J. Davidson. *The nature of emotion: Fundamental questions.* Oxford University Press, 1994.

[EE00] Ralf Erber and Maureen Wang Erber. The self-regulation of moods: Second thoughts on the importance of happiness in everyday life. *Psychological Inquiry*, 11(3):142–148, 2000.

[Ekm99] Paul Ekman. Basic emotions. *In Michael J. Power (Ed.), Handbook of cognition and emotion*, pages 45–60, 1999.

[Ell92] Clark Davidson Elliott. *The affective reasoner: a process model of emotions in a multi-agent system*. PhD thesis, Northwestern University, 1992.

[Ell93] David Ellis. A behavioural approach to information retrieval system design. *Journal of Documentation*, 45(3):171–212, 1993.

[Els04] Jon Elster. Emotions and rationality. *In A. S. R. Manstead & Nico H. Frijda Feelings (Eds.), Feelings and emotions: the Amsterdam symposium*, pages 30–48, 2004.

[EO79] Paul Ekman and Harriet Oster. Facial Expressions of Emotion. *Annual Reviews in Psychology*, 30(1):527–554, 1979.

[EP90] Norman S. Endler and James D. Parker. Multidimensional assessment of coping: A critical evaluation. *Journal of Personality and Social Psychology*, 58(5):844–854, 1990.

[ES87] Nancy Eisenberg and Janet Strayer. Critical issues in the study of empathy. empathy and its development. *In Nancy Eisenberg & Janet Strayer (Eds.), Empathy and its development*, 1987.

[ES06] Andrea Esuli and Fabrizio Sebastiani. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation*, volume 6, pages 417–422, 2006.

[ES09] Andrea Esuli and Fabrizio Sebastiani. Enhancing opinion extraction by automatically annotated lexical resources. In *Proceedings of the 4th Language Technology Conference*, pages 224–228, 2009.

[FEM05] Karen E. Fisher, Sandra Erdelez, and Lynne McKechnie. *Theories of information behavior*. Information Today Inc, 2005.

[Fes54] Leon Festinger. A theory of social comparison. *Human relations*, 7:117–140, 1954.

[Fes60] Leon Festinger. A theory of cognitive dissonance. *Journal of Consulting Psychology*, 24:349–354, 1960.

[FH85] Allan Fenigstein and Ronald G. Heyduk. Thought and action as determinants of media exposure. *In D. Zillmann & J. Bryant (Eds.), Selective exposure to communication*, pages 113–139, 1985.

[FP96] Tom Fawcett and Foster Provost. Combining data mining and machine learning for effective user profiling. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pages 8–13, 1996.

[Fri87] Nico H. Frijda. Emotion, cognitive structure, and action tendency. *Cognition & Emotion*, 1(2):115–143, 1987.

[Fri93] Nico H. Frijda. Moods, emotion episodes, and emotions. *In Michael Lewis & Jeannette M. Haviland (Eds.), Handbook of Emotions*, pages 381–403, 1993.

[Fri01] Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232, 2001.

[FRS64] Leon Festinger, Henry W Rieken, and Stanley Schachter. *When prophecy fails: A social and psychological study of a modern group that predicted the destruction of the world*. Harper and Row, 1964.

[GAT98] Laura Guerrero, Peter Andersen, and Melanie Trost. Communication and emotion: Basic concepts and approaches. *In Peter A. Andersen (Ed.), Handbook of communication and emotion: Research, theory, applications, and contexts*, pages 3–27, 1998.

[GH06] Ronald E. Goldsmith and David Horowitz. Measuring motivations for online opinion seeking. *Journal of Interactive Advertising*, 6(2):1–16, 2006.

[Gri89] Paul E. Griffiths. The degeneration of the cognitive theory of emotions. *Philosophical Psychology*, 2(3):297–313, 1989.

[GS09] Sreenivas Gollapudi and Aneesh Sharma. An axiomatic approach for result diversification. In *Proceedings of the 18th international conference on World wide web*, pages 381–390, 2009.

[GZ06] Andrew B. Goldberg and Xiaojin Zhu. Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization. In *Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing*, pages 45–52, 2006.

[HBW92] Micheline Hancock-Beaulieu and Stephen Walker. An evaluation of automatic query expansion in an online library catalogue. *Journal of Documentation*, 48(4):406–421, 1992.

[HCZ04] Zan Huang, Hsinchun Chen, and Daniel Zeng. Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. *ACM Transactions on Information Systems (TOIS)*, 22:116–142, January 2004.

[Her00] Jonathan Lee Herlocker. *Understanding and improving automated collaborative filtering systems*. PhD thesis, University of Minnesota, 2000.

[Hof84] Martin L. Hoffman. Interaction of affect and cognition in empathy. *In Carroll Ellis Izard & Jerome Kagan (Eds.), Emotions, cognition, and behavior*, pages 103–131, 1984.

[Hof03] Thomas Hofmann. Collaborative filtering via gaussian probabilistic latent semantic analysis. *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 259–266, 2003.

[Hof04] Thomas Hofmann. Latent semantic models for collaborative filtering. *ACM Transactions on Information Systems*, 22(1):89–115, 2004.

[IBM91] Carroll Ellis Izard, Edmund S. Bartlett, and Alan G. Marshall. *Patterns of emotions; a new analysis of anxiety and depression*. Plenum Publishing Corporation, 1991.

[Ing93] Peter Ingwersen. Cognitive perspectives of information retrieval interaction: elements of a cognitive ir theory. *Journal of Documentation*, 52(1):3–50, 1993.

[Ise84] Alice M. Isen. Toward understanding the role of affect in cognition. *In Robert S. Wyer and Thomas K. Srull (Eds.), Handbook of social cognition*, 3:179–236, 1984.

[Iza71] Carroll Ellis Izard. *The face of emotion.* Appleton-Century-Crofts, 1971.

[Iza91] Carroll Ellis Izard. *The Psychology of Emotions*. Plenum Publishing Corporation, 1991.

[Jam84] William James. What is an Emotion? *Mind*, 9(34):188–205, 1884.

[JK02] Kalervo Järvelin and Jaana Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems*, 20:422–446, 2002.

[Jos98] Joemon M. Jose. *An integrated approach for multimedia information retrieval*. PhD thesis, Robert Gordon University, 1998.

[JSBS98] Major Bernard J. Jansen, Amanda Spink, Judy Bateman, and Tefko Saracevic. Real life information retrieval: a study of user queries on the web. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 5–17, 1998.

[JWMG09] Niklas Jakob, Stefan Hagen Weber, Mark Christoph Müller, and Iryna Gurevych. Beyond the stars: exploiting free-text user reviews to improve the accuracy of movie recommendations. In *Proceeding of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, pages 57–64, 2009.

[Kan03] Hang-Bong Kang. Affective content detection using HMMs. In *Proceedings of the eleventh ACM international conference on Multimedia*, pages 259–262, 2003.

[KCSL05] Fang-Fei Kuo, Meng-Fen Chiang, Man-Kwan Shan, and Suh-Yin Lee. Emotion-based music recommendation by association discovery from film music. In *Proceedings of the 13th annual ACM international conference on Multimedia*, MULTIMEDIA ’05, pages 507–510, New York, NY, USA, 2005. ACM.

[KL10] Marian Kogler and Mathias Lux. Bag of visual words revisited: an exploratory study on robust image retrieval exploiting fuzzy codebooks. In *Proceedings of the Tenth International Workshop on Multimedia Data Mining*, MDMKDD '10, pages 3:1–3:6, New York, NY, USA, 2010. ACM.

[kLfClC06] Cane Wing ki Leung, Stephen Chi fai Chan, and Fu lai Chung. Integrating collaborative filtering and sentiment analysis: A rating inference approach. In *Proceedings of The ECAI 2006 Workshop on Recommender Systems*, pages 62–66, 2006.

[KM03] Dan Klein and Christopher D. Manning. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pages 423–430, 2003.

[Kno03] Silvia Knobloch. Mood adjustment via mass communication. *Journal of Communication*, 53(2):233–250, 2003.

[KRBH98] Joseph A. Konstan, John Riedl, AI Borchers, and Jonathan L. Herlocker. Recommender Systems: A GroupLens Perspective. In *Recommender Systems. Papers from 1998 Workshop. Technical Report WS-98-08*, pages 60–64, 1998.

[Krč08] Uroš Krčadinac. *Textual emotion recognition and creative visualization*. PhD thesis, University of Belgrade, 2008.

[Kuh91] Carol Collier Kuhlthau. Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42(5):361–371, 1991.

[Kuh93] Carol Collier Kuhlthau. A principle of uncertainty for information seeking. *Journal of Documentation*, 49(4):339–355, 1993.

[Kuh04] Carol Collier Kuhlthau. *Seeking meaning: a process approach to library and information services*. Libraries Unlimited, 2004.

[KVC10] Sunghwan Mac Kim, Alessandro Valitutti, and Rafael A. Calvo. Evaluation of unsupervised emotion models to textual affect recognition. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 62–70, 2010.

[Laz91] Richard S. Lazarus. *Emotion and Adaptation*. Oxford University Press, USA, 1991.

[LD92] Randy J. Larsen and Edward Diener. Promises and problems with the circumplex model of emotion. *Review of personality and social psychology*, 1992.

[LeD96] Joseph LeDoux. The emotional brain. *New York*, 1996.

[LEG01] Robert LaRose, Matthew S. Eastin, and Jennifer Gregg. Reformulating the Internet Paradox: Social Cognitive Explanations of Internet Use and Depression. *Journal of Online Behavior 1 (2)*, 2001.

[Lin64] Donald B. Lindsley. *Psychophysiology and Motivation*. Prentice-Hall, 1964.

[LL94] Richard S. Lazarus and Bernice N. Lazarus. *Passion and reason: Making sense of our emotions*. Oxford University Press, USA, 1994.

[LO92] Nancy L.Stein and Keith Oatley. Basic emotions: Theory and measurement. *Cognition & Emotion*, 1992.

[LS04] Hugo Liu and Push Singh. ConceptNet – Practical Commonsense Reasoning Tool-Kit. *BT Technology Journal*, 22(4):211–226, 2004.

[LSL03] Hugo Liu, Ted Selker, and Henry Lieberman. Visualizing the affective structure of a text document. *International Conference for Human-computer Interaction (CHI)*, pages 740–741, 2003.

[LYC07] Kevin Hsin-Yih Lin, Changhua Yang, and Hsin-Hsi Chen. What emotions do news articles trigger in their readers? In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 733–734, 2007.

[MAPJ09] Yashar Moshfeghi, Deepak Agarwal, Benjamin Piwowarski, and Jeomon M. Jose. Movie Recommender: Semantically Enriched Unified Relevance Model for Rating Prediction in Collaborative Filtering. In *Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, pages 54–65, 2009.

[Mar95] Gary Marchionini. *Information seeking in electronic environments.* Cambridge University Press, 1995.

[MC92] Marie-Louise Mares and Joanne Cantor. Elderly viewers' responses to televised portrayals of old age. *Communication Research*, 19(4):459–478, 1992.

[MCY08] Hemant Misra, Olivier Cappé, and François Yvon. Using lda to detect semantically incoherent documents. In *Proceeding of the 12th Conference on Computational Natural Language Learning*, pages 41–48, 2008.

[MD98] Leonard L. Martin and Blossom Davies. Beyond hedonism and associationism: A configural view of the role of affect in evaluation, processing, and self-regulation. *Motivation and Emotion*, 22:33–51, 1998.

[Mil95] George A. Miller. WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41, 1995.

[Mil98] John Stuart Mill. Utilitarianism. *Utilitarianism: and, On liberty*, pages 181–235, 1998.

[Miz96] Stefano Mizzaro. A cognitive analysis of information retrieval. In *Proceeding of the Second International Conference on Conceptions in Library and Information Science*, pages 233–250, 1996.

[Miz97] Stefano Mizzaro. Relevance: The whole history. *Journal of the American Society for Information Science*, 48(9):810–832, 1997.

[MJZ04] Bamshad Mobasher, Xin Jin, and Yanzan Zhou. Semantically Enhanced Collaborative Filtering on the Web. *Lecture Notes in Computer Science*, 3209:57–76, 2004.

[MMN02] Prem Melville, Raymod J. Mooney, and Ramadass Nagarajan. Content-boosted collaborative filtering for improved recommendations. In *Proceedings of the National Conference on Artificial Intelligence*, pages 187–192, 2002.

[MO06] Craig Macdonald and Iadh Ounis. The trec blogs06 collection: Creating and analysing a blog test collection. *Department of Computer Science, University of Glasgow Tech Report TR-2006-224*, 2006.

[Moo50] Calvin N Mooers. *The theory of digital handling of non-numerical information and its implications to machine economics*. Zator Co., 1950.

[MPI07] Shaikh Mostafa Al Masum, Helmut Prendinger, and Mitsuru Ishizuka. Emotion sensitive news agent: An approach towards user centric emotion sensing from the news. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, pages 614–620, 2007.

[MR00] Raymond J. Mooney and Loriene Roy. Content-based book recommending using learning for text categorization. In *Proceedings of the fifth ACM conference on Digital libraries*, pages 195–204, 2000.

[MRS08] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to information retrieval*. Cambridge University Press, 2008.

[MTL78] Robert McGill, John W. Tukey, and Wayne A. Larsen. Variations of box plots. *American Statistician*, 32(1):12–16, 1978.

[Nah04] Diane Nahl. Measuring the affective information environment of web searchers. *Proceedings of the American Society for Information Science and Technology*, 41(1):191–197, 2004.

[Nah05] Diane Nahl. Affective load. *Theories of information behavior*, pages 39–43, 2005.

[OCC90] Andrew Ortony, Gerald L. Clore, and Allan Collins. *The cognitive structure of emotions*. Cambridge University Press, 1990.

[Oli93] Mary Beth Oliver. Exploring the paradox of the enjoyment of sad films. *Human Communication Research*, 19(3):315–342, 1993.

[Oli03] Mary Beth Oliver. Mood management and Selective Exposure. *Communication and Emotion: Essays in honor of Dolf Zillmann*, pages 85–106, 2003.

[ONR05] Andrew Ortony, Donald A. Norman, and William Revelle. Affect and proto-affect in effective functioning. *Who Needs Emotions? The Brain Meets the Robot*, 2005.

[Ort08] Andrew Ortony. Affect and Emotions in Intelligent Agents: Why and How? *Affective Information Processing*, pages 11–21, 2008.

[PB97] Michael Pazzani and Daniel Billsus. Learning and Revising User Profiles: The Identification of Interesting Web Sites. *Machine Learning*, 27(3):313–331, 1997.

[PC09] Seung-Taek Park and Wei Chu. Pairwise preference regression for cold-start recommendation. In *Proceedings of the third ACM conference on Recommender systems*, pages 21–28, 2009.

[PHLG00] David M. Pennock, Eric Horvitz, Steve Lawrence, and C. Lee Giles. Collaborative filtering by personality diagnosis: A hybrid memory-and model-based approach. *Uncertainty in Artificial Intelligence*, pages 473–480, 2000.

[PK02] Sally Planalp and Karen Knie. Integrating verbal and nonverbal emotion (al) messages. *The verbal communication of emotions: interdisciplinary perspectives*, pages 55–79, 2002.

[PL05] Bo Pang and Lillian Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *Association for Computational Linguistics*, pages 115–124, 2005.

[PMN03] James W. Pennebaker, Matthias R. Mehl, , and Kate G. Niederhoffer. Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology*, pages 547–578, 2003.

[Rap06] David Rapaport. *Emotions and Memory*. Kessinger Publishing, 2006.

[RCH02] Bernard Rimé, Susanna Corsini, and Gwénola Herbette. Emotion, verbal expression, and the social sharing of emotion. *In S. R. Fussell (Eds.) The verbal communication of emotions: Interdisciplinary perspectives*, pages 185–208, 2002.

[RD06] Filip Radlinski and Susan Dumais. Improving personalized web search using result diversification. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 691–692, 2006.

[RIS+94] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the Conference on Computer Supported Cooperative Work*, pages 175–186, 1994.

[Rob77] Stephen E. Robertson. The Probability Ranking Principle in IR. *Journal of Documentation*, 33(4):294–304, 1977.

[RSLW06] Dawn T. Robinson, Lynn Smith-Lovin, and Allison K. Wisecup. Affect control theory. *In Jan E. Stets, Jonathan H. Turner. (Eds.) Handbook of the Sociology of Emotions*, pages 179–202, 2006.

[Rus80] James A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

[Rus89] James A. Russell. Measures of emotion. *In R. Plutchik & H. Kellerman (Eds.), Emoson: Theory research, and experience*, pages 83–111, 1989.

[Sar75] Tefko Saracevic. Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science*, 26(6):321–343, 1975.

[Sar07] Tefko Saracevic. Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology*, 58(13):1915–1933, 2007.

[SC88] Norbert Schwarz and Gerald L. Clore. How do i feel about it? the informative function of affective states. *Affect, cognition, and social behavior*, pages 44–62, 1988.

[Sch59] Ernest G. Schachtel. *Metamorphosis New York.* Basic Books, 1959.

[SCKP08] Mohammad Soleymani, Guillaume Chanel, Joep J.M. Kierkels, and Thierry Pun. Affective ranking of movie scenes using physiological signals and content analysis. In *Proceeding of the 2nd ACM workshop on Multimedia semantics*, pages 32–39, 2008.

[SDP09] Mohammad Soleymani, Jeremy Davis, and Thierry Pun. A collaborative personalized affective video retrieval system mohammad soleymani. *New York*, pages 0–1, 2009.

[SDSO68] Philip J. Stone, Dexter C. Dunphy, Marshall S. Smith, and Daniel M. Olgilvie. The general inquirer: A computer approach to content analysis. *Journal of Regional Science*, 8(1), 1968.

[Seb02] Fabrizio Sebastiani. Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1):1–47, 2002.

[See04] Sean Seepersad. Coping with loneliness: adolescent online and offline behavior. *CyberPsychology & Behavior*, 7(1):35–39, 2004.

[Sha08] Mostafa Al Masum Shaikh. *An analytical approach for affect sensing from text*. PhD thesis, University of Tokyo, 2008.

[SJRS07] Karen Spärck-Jones, Stephen E. Robertson, and Mark Sanderson. Ambiguous requests: implications for retrieval tests, systems and theories. *Proceedings of Special Interest Group on Information Retrieval Forum*, 41:8–17, 2007.

[SKKR01] Badrul Sarwar, George Karypis, Joseph Konstan, and John Reidl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295, 2001.

[SM86] Gerard Salton and Michael J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., 1986.

[SM07] Carlo Strapparava and Rada Mihalcea. Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 70–74, 2007.

[SMO10a] Rodrygo L.T. Santos, Craig Macdonald, and Iadh Ounis. Exploiting query reformulations for web search result diversification. In *Proceedings of the 19th international conference on World wide web*, pages 881–890, 2010.

[SMO10b] Rodrygo L.T. Santos, Craig Macdonald, and Iadh Ounis. Selectively diversifying web search results. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 1179–1188, 2010.

[Sol73] Robert C. Solomon. Emotions and choice. *The Review of Metaphysics*, pages 20–41, 1973.

[Sol77] Robert C. Solomon. The Rationality of the Emotions. *Southwestern (The) Journal of Philosophy Norman, Okla.*, 8(2):105–114, 1977.

[Sou90] Ronald De Sousa. *The rationality of emotion*. The MIT Press, 1990.

[SPI08] Mostafa Al Masum Shaikh, Helmut Prendinger, and Mitsuru Ishizuka. Sentiment assessment of text by analyzing linguistic features and contextual valence assignment. *Applied Artificial Intelligence*, 22(6):558–601, 2008.

[SPI09] Mostafa Al Masum Shaikh, Helmut Prendinger, and Mitsuru Ishizuka. A Linguistic Interpretation of the OCC Emotion Model for Affect Sensing from Text. *Affective Information Processing*, pages 45–73, 2009.

[SPMO10] Rodrygo Santos, Jie Peng, Craig Macdonald, and Iadh Ounis. Explicit search result diversification through sub-queries. In *Advances in Information Retrieval*, pages 87–99. Springer, 2010.

[SS62] Stanly Schachter and Jerome Singer. Cognitive, social, and physiological determinants of emotional state. *Psychological review*, 69(5):379–399, 1962.

[SS88] Jefferson A. Singer and Peter Salovey. Mood and memory: evaluating the network theory of affect. *Clinical Psychology Review*, 1988.

[SSKO87] Phillip Shaver, Judith Schwartz, Donald Kirson, and Cary O'Connor. Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, 52(6):1061–1086, 1987.

[Sul02] John R. Suler. Identity management in cyberspace. *Journal of Applied Psychoanalytic Studies*, 4(4):455–459, 2002.

[SV04] Carlo Strapparava and Alessandro Valitutti. WordNet-Affect: an affective extension of WordNet. In *Proceedings of the 4th International Conference on LREC*, pages 1083–1086, 2004.

[SVS06] Carlo Strapparava, Alessandro Valitutti, and Oliviero Stock. The affective weight of lexicon. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, pages 423–426, 2006.

[SVS07] Carlo Strapparava, Alessandro Valitutti, and Oliviero Stock. Dances with words. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1719–1724, 2007.

[Tom63] Silvan S. Tomkins. *Affect, imagery, consciousness*. Springer, 1963.

[UF98] Lyle H. Ungar and Dean P. Foster. Clustering methods for collaborative filtering. In *Proceedings of the Workshop on Recommendation Systems*, pages 112–125, 1998.

[Vor03] Peter Vorderer. Entertainment theory. *Communication and emotion: Essays in honor of Dolf Zillmann*, pages 131–153, 2003.

[vR79] C. J. van Rijsbergen. *Information Retrieval, 2nd Ed.* Butterworth, 1979.

[WdVR08] Jun Wang, Arjen P. de Vries, and Marcel J. T. Reinders. Unified relevance models for rating prediction in collaborative filtering. *ACM Transactions on Information Systems*, pages 1–42, 2008.

[Wil93] Theresa Ann Wilson. On user studies and information needs. *Journal of Documentation*, 37(1):3–15, 1993.

[Wil95] Raymond Williams. *The sociology of culture*. University of Chicago Press, 1995.

[Wis86] Lauren Wispé. The distinction between sympathy and empathy: To call forth a concept, a word is needed. *Journal of Personality and Social Psychology*, 50(2):314–321, 1986.

[WT10] Pinata Winoto and Tiffany Y. Tang. The role of user mood in movie recommendations. *Expert Systems with Applications*, 2010.

[WW11] Elizabeth Yost Hammer Wayne Weiten, Dana S. Dunn. *Psychology applied to modern life: Adjustment in the 21st century*. Wadsworth Pub Co, 2011.

[WWC05] Janyce Wiebe, Theresa Wilson, and Claire Cardie. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2):165–210, 2005.

[WZ09] Jun Wang and Jianhan Zhu. Portfolio theory of information retrieval. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 115–122, 2009.

[You61] Paul Thomas Young. *Motivation and emotion*. Wiley New York, 1961.

[YZLG09] Kai Yu, Shenghuo Zhu, John Lafferty, and Yihong Gong. Fast nonparametric matrix factorization for large-scale collaborative filtering. In *Proceedings of Special Interest Group on Information Retrieval*, pages 211–218, 2009.

[ZB85] Dolf Zillmann and Jennings Bryant. Affect, mood, and emotion as determinants of selective exposure. *In D. Zillmann & J. Bryant (Eds.), Selective exposure to communication*, pages 157–190, 1985.

[ZC77] Dolf Zillmann and Joanne R. Cantor. Affective responses to the emotions of a protagonist. *Journal of Experimental Social Psychology*, 13(2):155–165, 1977.

[ZCL03] Cheng Xiang Zhai, William W. Cohen, and John Lafferty. Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 10–17, 2003.

[Zha08] Cheng Xiang Zhai. Statistical language models for information retrieval a critical review. *Foundation and Trends in Information Retrieval*, 2:137–213, 2008.

[ZHB75] Dolf Zillmann, T. Alan Hay, and Jennings Bryant. The effect of suspense and its resolution on the appreciation of dramatic presentations. *Journal of Research in Personality*, 9(4):307–323, 1975.

[Zil80] Dolf Zillmann. Anatomy of suspense. *In P. H. Tannenbaum (Ed.), The Entertainment functions of television*, pages 291–301, 1980.

[Zil85] Dolf Zillmann. The experimental exploration of gratification from media entertainment. *In K. E. Rosengren, L. A. Wenner, & P. Palmgreen (Eds.), Media gratifications research: current perspectives*, pages 225–239, 1985.

[Zil88a] Dolf Zillmann. Mood management through communication choices. *American Behavioral Scientist*, 31(3):327–340, 1988.

[Zil88b] Dolf Zillmann. Mood management: Using entertainment to full advantage. *In L. Donohew, H.E. Sypher, & E.T. Higgins (Eds.), Communication, social cognition, and affect*, pages 147–171, 1988.

[Zil91] Dolf Zillmann. The logic of suspense and mystery. *In D. Zillmann & J. Bryant (Eds.), Responding to the screen: reception and reaction processes*, pages 281–303, 1991.

[Zil98] Dolf Zillmann. The psychology of the appeal of portrayals of violence. *In J. H. Goldstein (Ed.), Why we watch: the attractions of violent entertainment*, pages 179–211, 1998.

[Zil00a] Dolf Zillmann. Humor and comedy. *In D. Zillmann, &P. Vorderer (Eds.), Media entertainment: The psychology of its appeal, LEA's communication series*, pages 37–57, 2000.

[Zil00b] Dolf Zillmann. Mood management in the context of selective exposure theory. *In M. E. Roloff (Ed.), Communication Yearbook*, 23:103–123, 2000.

[ZW85] Dolf Zillmann and Jacob Wakshlag. Fear of victimization and the appeal of crime drama. *In D. Zillmann & J. Bryant (Eds.), Selective exposure to communication*, pages 141–156, 1985.