Social Contextuality and Conversational Recommender Systems

Eoin Hurrell, B.Sc. (hons)

A dissertation submitted in fulfilment of the requirements for the award of

Doctor of Philosophy (Ph.D.)

to the

Dublin City University
School of Computing

Supervisor: Prof. Alan F. Smeaton

January 21, 2013

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed:

ID No: 55377919

Date:

Abstract

As people continue to become more involved in both creating and consuming information, new interactive methods of retrieval are being developed. In this thesis we examine conversational approaches to recommendation, that is, the act of suggesting items to users based on the systems understanding of them. Conversational recommendation is a recent contribution to the task of information discovery. We propose a novel approach to conversation around recommendation, examining how it is improved to work with collaborative filtering, a common recommendation algorithm. In developing new ways to recommend information to people we also examine their methods of information seeking, exploring the role of conversational recommendation, using both interview and sensed brain signals.

We also look at the implications of the wealth of social and sensed information now available and how it improves the task of accurate recommendation. By allowing systems to better understand the connections between users and how their social impact can be tracked we show improved recommendation accuracy. We look at the social information around recommendations, proposing a directed influence approach between socially connected individuals, for the purpose of weighting recommendations with the wisdom of influencers. We then look at the semantic relationships that might seem to indicate wisdom (i.e. authors on a book-ranking site) to see if the "wisdom of the few" can be traced back to those conventionally considered wise in the area. Finally we look at "contextuality" (the ability of sets of contextual sensors to accurately recommend items across groups of people) in recommendation, showing that different users have very different uses for context within recommendation.

This thesis shows that conversational recommendation can be generalised to work well with collaborative filtering, that social influence contributes to recommendation accuracy, and that contextual factors should not be treated the same for each user.

Acknowledgements

Firstly I'd like to thank my supervisor, Alan Smeaton for all his guidance and support during my time producing this work, offering valuable advice at numerous crossroads. Thanks also to Science Foundation Ireland (SFI) and the Dublin City University Office of the Vice-President of Research, responsible for funding this research.

Thanks to all the current and past members of Clarity Centre for web sensor technology who offered advice and support. I would particularly like to thank Cathal Gurrin and Hyowon Lee, whose advice, feedback and collaboration were invaluable to me in my work. Many thanks to everyone who made PhD work an enjoyable experience.

I would also like to thank my friends; Marc, Colin, Graham, Gaëlle, Flash and Rob, for offering help and respite, and being understanding when I was occupied.

My parents, Maeve and Terry, deserve all the thanks I can give them for the years of providing the good home and environment that lead me to where I am today. Thanks also to my brother Cormac, who has in his own way always been there for me.

Finally I would like to thank my wonderful girlfriend Lisa, for her constant love, support and encouragement, even through the difficult process of producing this work.

Contents

Abstract **i**

Acknowledgements **ii**

List of Figures **vii**

List of Tables **ix**

List of Publications **xi**

1 Introduction **1**

1.1 Conversational Recommendation 2

1.2 Primary Hypothesis and Research Questions 4

1.3 Research Methodology 6

1.3.1 Metrics 8

1.4 Organisation of the Thesis 10

1.5 Origins of the Material 10

2 Related Work **11**

2.1 Recommender Systems 12

2.1.1 Collaborative Filtering 14

2.1.2 Content-based Recommendation 17

2.2 Conversational Recommendation 19

2.3 Social Recommendation 23

2.4 Contextual Recommendation 27

2.5 Summary 31

3 Conversational Recommendation 33

3.1 Introduction 33

3.2 Collaborative Conversational Recommendation 35

3.3 Design and System Outline 40

3.3.1 The Interaction Approach 40

3.3.2 The MovieQuiz Application 43

3.4 Evaluation 46

3.4.1 Interaction Analysis 46

3.4.2 Interactions As A Data Source For Improved Accuracy 49

3.4.3 Discussion 51

3.5 Comparison to Related Work 54

3.5.1 Collaborative Filtering and Conversation 54

4 Combined Recommendation in a Conversational Interface 56

4.1 Method 59

4.2 The Recommendation Architecture 64

4.2.1 Initial Recommendation 65

4.2.2 The Interactive Multimedia Component 65

4.3 Evaluation 69

4.3.1 User Survey 69

4.4 Discussion 74

4.5 Comparison to Related Work 76

4.5.1 Explanation And Knowledge 76

4.6 Chapter Conclusions and Answer to Research Question 79

5 Information Seeking 81

5.1 Information Seeking 81

5.2 User Study in Conversational Recommender Systems 83
5.2.1 Approach 83
5.2.2 Evaluation 84
5.2.3 Discussion 87
5.3 Neural Reactions to Recommended Items 88
5.3.1 Approach 92
5.3.2 Evaluation - EEG Analysis 94
5.3.3 Experimental Results 97
5.3.4 Discussion 99
5.4 Comparison to Related Work 101
5.4.1 Information Seeking in Conversational Recommendation . . . 101
5.4.2 EEG Experimentation and the Examination of Recommendation Response 102
5.5 Chapter Conclusions and Answer to Research Question 103

6 Social Context **105**

6.1 Social Context 105
6.2 Social Dataset 107
6.3 Social Trail 108
6.3.1 Examining Social Influences 110
6.3.2 Weighting Social Influences 112
6.3.3 Results 113
6.3.4 Discussion 119
6.4 Weighting Social influences 121
6.4.1 Approach 122
6.4.2 Results 125
6.4.3 Discussion 128
6.5 Comparison to Related Work 129

6.5.1 Social Trail 129
6.5.2 Expert Authority 131
6.6 Chapter Conclusions and Answer to Research Question 132

7 Context 134
7.1 Context and Recommendation 134
7.2 Shared or Sensed Context in Conversational Recommendation 135
7.2.1 Approach 137
7.2.2 Results 138
7.2.3 Discussion 139
7.3 Contextuality - Context Sets and their Usefulness 140
7.3.1 Approach 140
7.3.2 Results 144
7.3.3 Discussion 145
7.4 Comparison to Related Work 146
7.4.1 Views of Context 146
7.4.2 Contextual Feature Importance 148
7.5 Conclusion and Answer to Research Question 149

8 Conclusions 150
8.1 Answers to Research Questions 151
8.2 Recommendations Based on Work 153
8.3 Summary of Contributions 154
8.4 Future Directions 156

Appendix 159

Bibliography 163

List of Figures

2.1 The layers of information seeking. 20

3.1 The distribution of items in the MovieLens dataset when plotted using our measurements. 43

3.2 MovieLens collection dissected according to the user's choices. 45

3.3 The MovieQuiz application interface. 45

3.4 RMSE accuracy for the 4 CF CR approaches tested. 51

4.1 Exercise Builder interface, seeking to make route planning for exercise easier. 59

4.2 The red pin designates the route start, blue can be moved to modify the route. 66

4.3 Exercise Builder provides information about route difficulty through elevation. 67

4.4 Exercise Builder's embedded photos can be interacted with to see larger versions. 68

4.5 Participant experience levels. 71

4.6 Participant familiarity with test area on a scale of 1 (Not familiar) to 5 (Very familiar). 71

4.7 Participant rating of initial route quality on a scale of 1 (Poor) to 5 (Very good). 72

4.8 Rating distribution of Exercise Builder's multimedia usefulness. . . . 73

5.1 How often do you watch movies, either at home or in the cinema? . . 87

5.2 Would you consider yourself knowledgeable about movies? 87
5.3 How many of the movies in the system did you recognise? 88
5.4 What did you think of the quality of the movies suggested by the system ? 88
5.5 Did you feel the movie recommender offered a good selection of movies you otherwise wouldn't have heard of seen ? 89
5.6 What did you think of the interface ? 89
5.7 Do you think the interface help you find good films ? 90
5.8 How easy was it to state a preference between two movies in the movie quiz ? 90
5.9 Did you find using the interface preferable to just being given a list of suggestions ? 91
5.10 A typical EEG setup using 16 electrodes. In our work we use six. . . 92
5.11 An example recommendation from our brain dataset 94
5.12 EEG node placement diagram. 96
5.13 EEG difference between accept/reject signals in Recommended items. 97
5.14 EEG difference between accept/reject signals in Control items. 98
5.15 EEG difference between accept/reject signals in Interacted With items. 98

7.1 Tweet density over time, from public Dublin-based Twitter users over time 141

List of Tables

3.1 CF Conversational Recommender Interaction Analysis. 48

3.2 RMSE scores of relative preference session data integration. 50

4.1 Questions asked of users. 70

4.2 Runner familiarity with area and multimedia usefulness. 73

4.3 Exercise expertise and multimedia usefulness. 73

5.1 MovieQuiz User Survey Questions 86

5.2 User responses to control route image stimulus 100

5.3 User responses to recommended route image stimulus 100

5.4 User responses to interacted recommended route image stimulus . . . 100

6.1 Rating statistics for the Goodreads dataset. 108

6.2 Miscellaneous statistics for the Goodreads dataset 108

6.3 Example Goodreads rating details 109

6.4 Rating Sentiment Influence on Goodreads dataset 116

6.5 Rating Influence on Reading and Rating in Goodreads dataset 116

6.6 RMSE Accuracy of Socially-Aware Recommender Algorithm 117

6.7 Recall of Socially-Aware Recommender Algorithm 118

6.8 Area Under Curve (ROC) of Socially-Aware Recommender Algorithm 118

6.9 Comparison of Socially-Aware Recommender tests (20% test) 118

6.10 RMSE values of Social-Role-Aware Recommender Algorithm 126

6.11 Area Under Curve (ROC) values of Social-Role-Aware Recommender Algorithm 126

6.12 Precision of Social-Role-Aware Recommender Algorithm 127
6.13 P@5 of Social-Role-Aware Recommender Algorithm 127
6.14 P@10 of Social-Role-Aware Recommender Algorithm 127
6.15 Recall of Social-Role-Aware Recommender Algorithm 127
6.16 Convergent vs Divergent Authors Read (RMSE) 128
6.17 Convergent vs Divergent Authors Similar (RMSE) 128

7.1 Survey questions 137
7.2 Context statistics 138
7.3 Descriptions of Dynamically Generated Features 142
7.4 Text features and the number of categories for each 143
7.5 The top average important features in deciding whether a user follows another 144
7.6 The most selected features by SVMs trained on individual users . . . 145

I Convergent vs Divergent Authors Read (P@5) 159
II Convergent vs Divergent Authors Similar (P@5) 159
III Convergent vs Divergent Authors Read (P@10) 159
IV Convergent vs Divergent Authors Similar (P@10) 160
V Convergent vs Divergent Authors Read (Precision) 160
VI Convergent vs Divergent Authors Similar (Precision) 160
VII Convergent vs Divergent Authors Read (Recall) 160
VIII Convergent vs Divergent Authors Similar (Recall) 160
IX Twitter features selected for context (part 1) 161
X Twitter features selected for context (part 2) 162

List of Publications

- E. Hurrell, A. F. Smeaton, and B. Smyth. Interactivity and Multimedia in Case-Based Recommendation.*FLAIRS Conference. AAAI Press, 2012. ISBN 978-1-57735-558-8*

- E. Hurrell and A. Smeaton. An Examination of User-Focused Context-Gathering Techniques in Recommendation Interfaces. *Irish Human Computer Interaction 2012, Galway, Ireland, 2012*

- E. Hurrell and A. Smeaton. Energy Saving Using Location Aware Sensor Networks. *Telecommunications Energy Conference (INTELEC), 2011 IEEE 33rd International, pages 14. IEEE, 2011a*

- E. Hurrell and A. F. Smeaton. The Benefits of Opening Recommendation to Human Interaction. *BCS HCI, pages 479484. ACM, 2011b*

- D. Orpen, C. Fay, D. Maher, T. Phelan, E. Hurrell, C. Foley, A. Smeaton, and D. Diamond. Remote Monitoring of Landfill Gases from Solid Waste Landfill (Including Real Time Data Integration to a Web Based Data Portal). 2011a

- D. Orpen, C. Fay, T. Phelan, E. Hurrell, C. Foley, A. Smeaton, and D. Diamond. Remote Monitoring of Landfill Gases from Solid Waste Landfill. 2011b

Chapter 1

Introduction

Information seeking is a task that has defined much of the modern Internet, from search to shopping online, with good reason. As the Internet expands the task of finding relevant information becomes increasingly non-trivial. There are many different contexts within which people may seek information, and their intent can be complex, vaguely expressed and difficult to determine. Part of this complexity is due to the fact that people may know what they like when they see it, but not how to search for it. Recommendation involves finding items that users might like based on what is understood of their interests, and is the task associated with predicting how someone will react to items in a collection in order to suggest the best items for them. It can however be a difficult task to determine if the results retrieved are optimal, as is the case with recommendation, where the user can only indicate degree of success tangentially after exploring.

There are also growing social factors around items on the Internet, with discussion and friendship forming around recommended items, and people sharing their opinions with others of products and services. People are now using devices with far more sensors than ever before to create an online presence that they use to discuss items they would have merely rated before. It is in this environment that conversational recommenders emerge; recommendations are turned into an active process of providing feedback and expressing relevance directly to a system so it may better recommend items.

This thesis examines conversational recommenders and their place with respect to the growing presence of social and contextual data available. The relationship between these three elements forms the core of our work as we look at whether all recommenders can be used in conversational recommendation, whether there is a place for social information and whether people have a unique approach to the context others display to them. The aim is to understand and support new forms of recommendation developed from this work.

This chapter will outline conversational recommendation in Section 1.1, after which we will describe the scope of this thesis and the aims, leading to the formulation of the primary hypothesis and research questions in Section 1.2. In Section 1.3 we describe the research methodology, with the structure of the thesis provided in Section 1.4. Finally we briefly note the origin of the material in Section 1.5.

1.1 Conversational Recommendation

Traditional recommendation systems such as the one found on Netflix [1] have been shifting to a more conversational approach in order to ease users into the recommendation process and acquire enough information from them in order to accurately recommend items. Recommendation, since it is not based on queries, uses past information about a person to suggest items they may like. It does this by collecting ratings or implicit feedback such as views and using them to build an understanding of the user from their actions.

Conversational recommendation however is a relatively new approach to recommendation; rather than fully relying on the information collected prior to a recommendation, conversational recommenders begin a conversation with "this is what

[1] `http://www.netflix.com`

we think you'll like" and allow people to correct, alter or otherwise explore the possibilities from that starting point. So far this conversation has been focused on metadata, such as providing feedback on the desired aperture for a camera. Empowering users to directly instruct a system that is in effect making suggestions for them is a useful way to both improve recommendation and gain more information about a user. This personalised information-filtering therefore relies on both well-described similar items to recommend and an understanding of the area of recommendation by the user.

Also becoming widely available is a plethora of social and contextual data about people and their relationships that remain unexplored for conversational recommenders. Websites like Goodreads[2], a site for people to share opinions on books with friends, provide a wealth of new data. Goodreads itself recently recorded its six millionth user registered and 200 millionth book catalogued [3]. This offers a wealth of data, not only the ratings of six million users over 200 million items to offer suggestions, but potentially the friendships between these users as well.

There is currently no single coherent attempt to benchmark the performance of systems designed to take into account new social and contextual sources while working without traditional metadata restrictions. In this research we wish to

1. Demonstrate a methodology that allows for discussion-mimicking item suggestions in content-free modalities.

2. Allow for the further in-depth study of information seeking behaviour using recorded data.

3. Use contextually-detected social interactions to enhance a system's mimesis and increase recommendation accuracy.

4. Examine the role of context for users in recommender systems.

[2]`http://www.goodreads.com`

[3]`http://www.goodreads.com/blog/show/307-goodreads-records-6-\ millionth-member-and-200-millionth-book-catalogued`

1.2 Primary Hypothesis and Research Questions

Having outlined what conversational recommenders are, we now state the primary hypothesis of this work.

Primary Hypothesis Conversational recommenders show great potential to be useful in offering in-situ suggestions and information seeking, but can be made more powerful by harnessing a user's social context.

In this thesis we will explore the ways in which we proved our primary hypothesis. We will begin by outlining the research questions that we derived from the hypothesis in order to prove it incrementally.

RQ1 How can we create conversational recommenders without intrinsic item knowledge?

Current methods of conversation such as critiquing are designed to leverage the metadata frequently found in content-based recommenders. However collaborative filtering for recommendation can effectively recommend items with no knowledge of what they are or any metadata, and as such seems incompatible with interaction without some degree of hybridisation. Here we investigate the possibilities of building systems that are interactive without using metadata in order to ensure that the principals of interactivity can be generalised across all recommendation methodologies. This will show conversational recommendation's general suitability to offering item suggestions as per our hypothesis, and not merely limited to similar content that is well-described. It will also allow for additional future work in collaborative filtering to involve interaction and compare traditional and conversational methods.

RQ2 Do conversational recommenders help fulfil a browsing information need?

Traditional recommenders have been studied from a human-recommender interface perspective and as part of information retrieval systems in information seeking, but conversational recommendation systems have yet to be studied in these ways. The interaction provided by conversational recommenders adds an entirely new dimension to these models. How users exercises their agency in these systems has been largely unaccounted for, as a user's traditional interaction was abstract from their information need, i.e. they shared their ratings of items with other users and then later expected the system to fulfil an information need because it "knows them". Key to this approach will be user involvement and browsing, hopefully improving the usefulness of recommendation in the so-called "just browsing" task it has historically claimed to be well-suited to. When given the ability to refine recommendations directly we wish to explore how this helps the user find acceptable items, as well as possible sensors that could further assist. One key element of this will explore whether a user's brain activity indicates measurable acceptance of one recommendation above another. User studies will also question users on how they used conversational systems and how they felt the systems enabled them to browse, along with system logs. This will confirm conversational recommendation's use in offering suggestions in a useful way to people.

RQ3 Can social relationships inferred from contextual cues prove useful in improving recommendation accuracy?

Next we follow the social activity of conversation with a social source of information. Given that recommenders suffer from well-recognised issues, including cold-start and rating noise problems, can social information, specifically what a user thinks of the opinions of friends and experts, help to produce more accurate item suggestions? Some research into this area has already been done, such as for basic groupings by Liu and Lee (2010), but in this work we will infer complex social relationships in order to produce a clearer picture of a user's social standing and its effect on recommendation. We will approach this problem on two fronts, using additional social information to refine rating accuracy using socially-driven weighting. This usage of social context will be distinct from social network recommenders, in that it will recommend with the aid of social networks, and not with a dependence on them. This will harness a person's social context to determine if it can be used as a data source in recommendation, improving the performance of the algorithmic side of conversational recommendation (which was the focus of RQ1).

RQ4 Can sensed or shared context be used to discover the unique criteria for any person's contextual recommendation?

Finally we will explore another dimension of context that may affect social context and its ability to effectively predict items. It has been traditional when using additional contextual sources (sensors such as GPS or offered data such as age) to use all available data or to design the usage around the task. We here postulate that there may in fact be a third strategy in context selection, designing around how individuals make use of context. The hope is that this will result in a dynamic approach to contextual recommendation that is free from any design-time bias and allows us to study the circumstances under which a selection of sensor usages are apt in the recommender domain. This will provide us with a study of the ways in which contexts such as social context information can be integrated into conversational recommenders.

1.3 Research Methodology

This research will focus on exploring ways to improve the quality of in-situ recommendations by mimicking the interactions with other people that usually generate them. We will begin by examining how people see recommendations through EEG analysis, and follow this with more complex experiments to test the idea of conversational contextual recommendation and user reaction to it. As the large-scale approaches of grouping people by shared interest or offering items similar to known good items are the standard, these will be used as a basis with a view to enriching them for the individual user by providing a more personal and exploratory method. This more personal conversational paradigm will be tested to see if it benefits from being altered based on the user's current position (either exact or abstract) and social standing (influence and view of the abilities of friends to recommend items).

This alteration will be explored in different systems, involving both intelligently altering the interface based on context and dynamically altering contextual criteria. This intelligently designed interface will dynamically change its line of questioning based on factors that would affect an ordinary conversation. Dynamic criteria alteration will examine how to encapsulate the idea that different sensed contexts matter in different ways, depending on those contexts. It is hoped these methods will help the user engage with the process so that more information can be inferred and users can be grouped more usefully. As an example, in a current commercial system a user not choosing to click on an item means nothing, the recommendation is not the focus of their usage. In an interactive recommendation system however where the recommended items are the focus it seems intuitive that if an item is ignored it is ignored for a reason.

An initial EEG experiment will be conducted to explore the effects of recommendation and interactive recommendation on brainwaves. Two systems will be developed in order to evaluate the strength of socially-relevant contextual cues in conversational recommendation. The first will explore the dynamic alteration of sensed-data usage in order to intelligently optimise the power of context in recommendation. The second will evaluate intelligent modification of a conversational flow in order to build a socially-relevant interactional context to improve recommendation accuracy. Both of these systems will be examined in terms of both recommendation-improving ability and user interaction, which will hopefully lead to methods to recommend an interface based on context. Further to this, user studies and system logs will explore the information seeking behaviour of users.

1.3.1 Metrics

Here we will describe the main metrics we will use to evaluate the effectiveness of our approaches.

Root Mean Square Error Root Mean Square Error is one of the most commonly used measures of accuracy in recommender systems. It measures the predictive ability of an algorithm in terms of the mistakes it makes. This is done by training the algorithm on some portion of the available ratings information and then checking the predicted ratings of the rest of the known ratings. A lower number corresponds to a more accurate predictive algorithm. Recently questions have been raised as to the usefulness of RMSE as a measure of recommender accuracy, as it treats rating numbers as interval data when it is ordinal data[4] (Koren and Sill (2011)). It is however the current scientific standard and as such we will use it to allow comparison of our work with other methods. Where we use RMSE we have tried to include other appropriate metrics as well to throughly test approaches.

Receiver Operating Characteristic (Area Under Curve) The Receiver Operating Characteristic (ROC) curve is a representation of the fraction of true positive items (in our case correctly recommended items a person has rated) to false positives (recommended items the user has not rated) at various thresholds (ratios of training to testing data). The area under this curve (AUC) is a single number used to compare algorithms. Area under the curve is widely used in machine learning and other fields along with recommendation. Recently however it has been suggested that it is a noisy measurement (Hanczar *et al.* (2010); Lobo *et al.* (2008)). While this inaccuracy is to be noted, it is still a standard test that we employ in order to be able to compare to existing

[4] `http://technocalifornia.blogspot.ie/2011/04/recommender-systems-were-doing-it-all.html`

approaches. Where we use AUC we have tried to include other appropriate metrics as well to throughly test approaches.

Precision Precision is a common metric in information retrieval to measure the effectiveness of algorithms. It is the measure of the fraction of retrieved items that are relevant. This is usually tested by withholding a number of user-rated items (i.e. known relevant items) from the system and seeing how many said system can find. Precision can be seen as the number of true positives in a collection divided by the number of items labelled as positive (including false positives). A higher precision is desirable in a system because it means the algorithm finds more relevant items to suggest than irrelevant, which will help the user find good items.

Recall Recall is another common information retrieval metric. It is the measure of the fraction of relevant items that are retrieved. Tested in the same way as precision, it can be seen as how many of the possible relevant results the algorithm returned. A higher recall is better, as it means the algorithm was better at finding all the good items in the collection.

Precision @ N While precision is a desirable quality since search and recommender algorithms frequently optimise for a short list of top results, overall precision might be a misleading metric. Precision @ N measures the precision of the first N items in an ordered list, effectively telling us how accurate the top N (N is usually 5 or 10) list is at providing good items. This in turn offers insight into how the algorithm will perform in situations where it is producing a short list of items, similar to a web search or "customers who bought this also bought" dialogue.

Other methods The above methods are the standard metrics used to test the effectiveness of recommender systems, however we have also employed other forms of analysis over the course of our work when deemed necessary, such as in cases where our approach made no changes to the underlying recommendation algorithm. In these cases we have described the metrics we used to test the approach in detail.

1.4 Organisation of the Thesis

This thesis will begin by outlining the currently existing work that relates to our research in Chapter 2. In Chapter 3 we look at how conversational recommenders can be implemented in environments without metadata (RQ1). Having looked at the algorithms and interfaces we then (in Chapter 5) look at how people can seek information in these conversational recommenders, including exploratory EEG experiments (RQ2). Having explored how conversation works we turn to social sources of information for Chapter 6, examining friendships and other relationships to see if we can infer a social impact that would contribute to recommendation accuracy (RQ3). Chapter 7 then looks at user motivations behind how they knowingly or unknowingly make use of context in their decisions to see if it may have an effect on how recommenders use context (RQ4). Finally we will conclude with a summary and look at how in answering our research questions we have formed a contribution to the field.

1.5 Origins of the Material

This thesis is partly based on papers that have already been published or have been submitted for publication. Early versions of the approach to interactive recommendation in Section 4 were detailed in Hurrell and Smeaton (2011), and later expanded to examine the effects in Hurrell *et al.* (2012). Finally our work in Section 7.2 is based on earlier work detailed in Hurrell and Smeaton (2012).

Chapter 2

Related Work

The work presented in this thesis is novel in two ways, firstly in its application and exploration of conversational recommendation and secondly in its exploration and incorporation of emerging real-time data sources. This chapter serves as a general introduction to literature relevant to this, covering areas intersecting with the work presented in this thesis. We will further discuss in each chapter how that chapter's work builds on and compares to the related work, here however we discuss the literature that forms the basis for the fields we are contributing to in this thesis.

We begin by discussing recommender systems in Section 2.1: we give a brief history of the field covering popular algorithms and uses, finishing with problems that these approaches currently suffer from. Next in Section 2.2 we take a more detailed look at conversational recommendation, the newer field growing from recommender systems. In Section 2.3 we examine the growing exploration of social sources of data, from such post-Web 2.0 sources as social networking websites or services that allow user annotation or creation of data (so-called "folksonomies"). We establish the current state of the art and issues faced in using social information. Lastly we explore the idea of context as detected from a number of sources including physical sensors, it's history and purpose, along with problems facing the field currently in Section 2.4.

2.1 Recommender Systems

Information Retrieval (IR) is the field of study which lead to the development of recommender systems. IR research has been conducted since the 1950s, studying among other things the ways people could retrieve needed information from document collections. At the time the most obvious application was library search; librarians looking for books or catalogue information, but as the Internet emerged and time passed the task of information retrieval moved from the smaller document collections of libraries to the exponentially larger collections of the World Wide Web. This meant retrieval went from an expert task to a mundane one, as people today commonly use robust search engines, but it also meant that new tasks became apparent. In search, the user is focused on finding one or more 'correct' answers to their queries, whereas on the Internet people are aware that there exist many things that they might like to see, but aren't searching for directly, either because they don't know about the specific item or don't know how to find it. Recommendation, as discussed in literature such as work by Resnick and Varian (1997) (or more recently by Ricci *et al.* (2011)), is the task of showing people items in a collection that they haven't seen but would like to see. One overview of the current state of IR has been written by Manning *et al.* (2008).

Information Seeking (IS) examines how people fulfil their information needs, whether through the use of IR systems or otherwise (discussed in depth by Marchionini (1995)). It has long been held that so-called "information search behaviour" a subset of information seeking behaviour (Wilson (1999b)), is not limited to search as the exclusive method of finding relevant information for users. Fallows and Project (2008) found that search engines are by far the most popular way to satisfy information needs online and Kuhlthau (1991) examined in detail how search is conducted by people on the web. Worth noting is the existence of "exploratory search", examined by Marchionini (2006b), using search engines in order to gain enough knowledge to understand how to search for the desired information, showing people who don't have a clear sense of what they want are actually looking for a means to find it. While many commercial or public recommender systems are supplementary to catalogues (with recommendations appearing in newsletters or, as on Amazon[5] in a "customers who bought this also bought" box) recommendation has been identified (by Belkin (2000)) as helping people understand the area so they can formulate coherent queries in search.

Another field similar to recommendation, information filtering, is worth noting as it attempts to match a user's interests to items using textual analysis and show them only things that are relevant, for example news articles. This can be seen as removing items considered irrelevant, in contrast to recommendation, which highlights items that are relevant (as discussed by Hanani *et al.* (2001)).

Since the 1990s recommenders have developed significantly, and many different algorithms exist to recommend items in different domains. Presently recommendation is almost as ubiquitous as search through its widespread uptake by businesses on the Internet and covering all kinds of services and products. Herlocker *et al.* (2004) discuss at length both the variety of uses of recommenders, from "Find Good Items" to "Just Browsing" (mentioned in more detail in Chapter 3), as well as metrics (including those mentioned earlier in Section 1.3.1) to evaluate the success of a system for a task, though new evaluations are constantly being suggested, including by Meyer *et al.* (2012).

Recommender systems are commonplace as a method for highlighting to consumers new items (recommendations can have significant impact in directing consumer behaviour as shown by Zhang, Jingjing (2011)), such as books, movies, websites, hotels or businesses, that will most probably be of interest or of use to them. This versatility and popularity has also been seen in the variety of research conducted on recommender systems, looking at music (Celma (2007)), medicine (Miyo

[5] `http://www.amazon.com`

et al. (2007)), traffic routes (Haigh *et al.* (1997)), known experts (McDonald and Ackerman (2000b)) and with high-profile competitions such as the million-dollar Netflix Prize[6]. Further examinations of recommender domain applications have been conducted by Montaner *et al.* (2003). In the following sections we will talk about the two major forms of recommender algorithm, as outlined by Adomavicius and Tuzhilin (2005).

2.1.1 Collaborative Filtering

Collaborative filtering (CF) is the most popularly implemented variant of recommendation algorithm at present, as discussed by Schafer *et al.* (2007). CF mimics "word-of-mouth" recommendations by making connections between items and people based on ratings, to offer 'serendipitous' item discovery, the things people didn't know they wanted. Offering items based on the ratings of like-minded users began with the Tapestry project (Goldberg *et al.* (1992)), which allowed users to annotate and filter documents by the people who made annotations to them. This allowed for a system of personally identifying people whose opinions were trusted in order to see what they thought was relevant, a process which has since been engrained in algorithm. Research by Resnick *et al.* (1994) followed by Shardanand and Maes (1995) identified the potential of calculating the similarity between users based on the common ratings they provided and using that information to predict probable interest in new items. Breese *et al.* (1998) performed some the first large-scale work on CF recommenders, evaluating and optimising then-current research. More recently comprehensive guidelines for the evaluation of such recommender systems have been outlined (by Herlocker *et al.* (2004)), which have become the standard practices we follow here.

CF relies on some form of indicator of opinion in order to connect people and offer items. It is common to see a rating system, between one and five stars, or like/dislike

[6] `http://www.netflixprize.com/`

(in situations where extreme opinions are more standard, such as YouTube [7]), but increasing study has delved into implicit feedback (Douglas and Jinmook (1998)) with some signals such as views or (for music) plays providing good indications of interest (Celma (2007)). All these forms of rating the personal relevance of items can contribute to how CF algorithms find a person's peers, predict what they will think of items and make recommendations. A user-item rating matrix can be formed from this activity, showing the ratings a user gave each item, with blank spaces where no rating was given (in practical examples most of the cells in these matrices are blank, as it is rare for a user to rate everything). The task of a CF algorithm can be seen as filling in these blanks with predicted values.

There are two currently accepted methods of CF, memory-based or model-based. Both are grounded in machine learning, *model-based collaborative filtering* takes the rating information of a user and has a training phase in which it constructs and trains a predictive model from those ratings, using machine learning algorithms such as Naive Bayes (Breese *et al.* (1998)). Creating a list of item recommendations is then a case of applying the model. This approach is said to be *eager* because all of the computational work is performed immediately, rather than when a user's recommendations are requested.

Memory-based collaborative filtering defers the computation of a user's recommendations until it is requested, and for this reason is sometimes known as *lazy* recommendation. It is memory-based because it stores all user ratings in memory. It uses the k-Nearest Neighbour algorithm, a machine learning algorithm, in either *user-based filtering* or *item-based filtering*. Similarity in both cases is computed using Pearson correlation or Cosine similarity (discussed further by Breese *et al.* (1998)). User-based filtering finds people who are similar to, through a history of agreeing with, the current user. This places the user in a 'neighbourhood' of their similar peers. All the items that their peers have seen but they have not are consid-

[7] `http://youtube-global.blogspot.ie/2009/09/five-stars-dominate-ratings.html`

ered as possible recommendations. These potential items are sorted based on their occurrences in the neighbourhood, with a weighted aggregate of these numbers used to generate the recommendations (Herlocker *et al.* (1999)). This can be seen as showing the user the most popular items among their peers. Item-based filtering (described by Sarwar *et al.* (2001) and used by Amazon[8], Linden *et al.* (2003)) computes the similarity between items a person likes and other items. Items are said to be similar if the same users rated them in the same way. If an item is similar to multiple items the user has rated its similarity score is the sum of all similarity scores it has. The recommendation list comes from a sorted aggregate list of these similarity scores. These similarities, for both user and item-based filtering, can be precomputed as a batch job to offer more efficient memory-based recommendations, in the style of eager model-based ones, at the cost of missing new ratings provided since the computation took place. The current benchmark for CF, that is the say the optimal implementation of it, is matrix factorisation as explored by Koren *et al.* (2009) and popularised by the Netflix prize. This approach characterises both users and items as vectors of factors, high correspondence between factors results in recommendations.

The flexibility and relatively low information requirement of CF (only needing user ratings, as compared to content-based approaches which as we will discuss need more up-front information) has contributed to its ubiquity in recommendation. A great benefit of the CF approach is that it sees items as black boxes with no attached metadata, ensuring it generalises to any task, and between completely different items, for example content-based methods will only recommend books to people who rate books. It is also easy to extend with new data sources or implicit feedback mechanisms, such as tagging systems, as investigated by Tso-Sutter *et al.* (2008).

There are of course ongoing problems and research questions involved in CF. Firstly, the system requires a non-trivial amount of information about a user be-

[8] `http://www.amazon.com`

fore its recommendations can approach a reliable accuracy and offer the user good suggestions. This is known as the "cold start" problem and pure CF algorithms continue to face it, which may somewhat explain why recommendations are frequently relegated to the sidebar of websites etc. Secondly, and again related to rating sparsity is what has been termed "the long tail" (Brynjolfsson *et al.* (2006)). The long tail is the collection of items that have very few ratings in a system, which results in them being less-recommended, regardless of suitability. Work is being done to address this, such as by Park and Tuzhilin (2008) but it is an ongoing area of study. Methods to both get feedback and potentially allow users to find 'long-tail' items are outstanding issues we investigate in this work.

2.1.2 Content-based Recommendation

Content-based recommendation is another class of recommendation that does not rely on word of mouth but rather the content of items to be recommended (Pazzani and Billsus (2007)). Items in content-based recommendation are well-described by metadata, so recommendable books have attached information like "author", "publisher", "genre" and so forth, and recommendation becomes the task of finding items with similar attributes. This has accordingly lead to much research into the best ways to harness such attributes (such as by Tso and Schmidt-Thieme (2006)). Case-based recommendation is a type of content-based recommender concerned with the re-use of well-described 'cases', situations that offered good solutions for the recommendation scenario currently faced. It has been explored by de Mántaras *et al.* (2005); Bridge *et al.* (2005) and more recently by Smyth (2007), with such useful applications as recommendation to groups (Jameson and Smyth (2007)) and route planning (McGinty and Smyth (2001, 2002a)). It is of importance here due to its extensive use in conversational recommender systems, which we will discuss in Section 2.2, such as work done by McCarthy *et al.* (2004b), showing the direction conversational recommenders have taken so far.

Content-based recommendation has many benefits. With collections that are already described by metadata it is easy to implement. It does not suffer from the cold-start problem of collaborative filtering, not relying on dense rating information but on attributes leads to a wealth of potential recommendations. Hybrids of the two approaches exist, as studied by Burke (2002b), as well as further offshoots of content-usage like knowledge-based recommendation (R. Burke (2000)), but these all suffer from placing an additional burden of metadata on the collection. As we will see, metadata is important for many conversational approaches but it limits recommendations to comparable items, if a user has only rated books then they can only be recommended items with "author" or similar attributes. For this reason it is a highly situational approach, and not used by e-commerce websites such as Amazon[9].

In summary, much work has been done on recommender systems and they continue to play a vital part in modern IR. However the dual problems of information overload and finding relevant information (explored by Anand and Mobasher (2005)) continue as the Internet grows. These issues, discussed as an IR problem by Belkin and Croft (1992) and a recommender problem specifically by Borchers *et al.* (1998), are convoluted by the growing using of smartphones for IR, as noted by Google *et al.* (2010), and the recently discussed problem that users think of ratings as ordinal ("This is two stars better than that"), rather than as intervals, as recommendation metrics assume (Koren and Sill (2011)). This has lead to, among other research, extensive studies of the new sensors that are now available (including by Smyth (2009)).

As we have mentioned at the start of this section, IS research tells us that users who are unsure what they are looking for are using exploratory search, or even recommenders themselves, in order to gain knowledge needed to fulfil their

[9] `http://www.amazon.com`

information needs. Clearly users wish to exercise their agency in some way when they are unsure what they want, since recommenders are in these cases not being used to fulfil information needs directly. This is compounded by the difficulty in studying recommenders within IS, as users have no real way to *seek* knowledge, except through tangental expressions of how well they liked other things, and the stated tasks suited to recommenders can be vaguely defined, such as "Just Browsing".

2.2 Conversational Recommendation

As we have seen, recommenders are designed to offer things without an explicit query, to learn based on observing. This is a rather passive form of information retrieval, and as has been shown by Marchionini (2006b) users who have an information need but can't form an explicit query still try to interact. Conversational (or sometimes interactive) recommenders were born out of explorations into what this apparent desire for interaction could offer and developed from relevance feedback (including work done by Salton and Buckley (1990)), an approach to IR that allows users to comment on whether items are relevant. This has extended in multiple interesting directions to make for more complex interactions, for example Campbell (2000). Work in information seeking by Wilson (1999b) discusses the layers of information seeking behaviour (Figure 2.1). These layers divide the search task and the actual act of searching, which in a recommendation context is aided by conversation.

Conversation offers recommenders a way to collect non-committal explicit interest, as recent work by Sparling and Sen (2011) has shown rating is sometimes difficult. Further it has been shown (Rafter and Smyth (2005)) that conversation shows the difference between *immediate* interests and *ongoing* likes and dislikes, important information that traditional recommendation has no way of capturing, though some research has occurred, such as work by Schafer *et al.* (2002). Beyond these algorithmic concerns Sinha *et al.* (2001) investigated the human computer in-

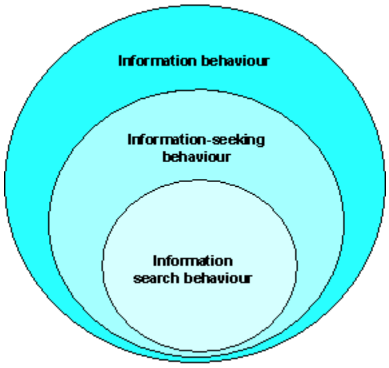

Figure 2.1: The layers of information seeking.

teraction perspective of recommendation, finding that "the ultimate effectiveness of an RS is dependent on factors that go beyond the quality of the algorithm" and that effective recommenders are seen to provide ways to refine recommendations in return for more effective recommendations.

Work by Ruthven (2008) has discussed the need for study into which IR tasks would benefit from interaction. Shimazu (2001) provided early work in conversational recommendation, proposing a novel approach to the interaction process; the "ExpertClerk" system asked about or proposed items to the user to provoke feedback. This has lead to work in preference elicitation (Knijnenburg and Willemsen (2009)) that seeks to make the process of getting enough information, to produce accurate recommendations, efficient. Conversation further developed with the idea of 'critiquing', Burke (2002a) did early work in this method of essentially correcting recommendations. In critique-based systems users restrict the size of the recommendable collection by providing feedback like "too expensive" on recommendations. This would restrict items based on their "price" attribute, and offer better recommendations as a result. This limits the approach to content-based or heavily metadata-enriched items and limits it to a subset of recommendation algorithms, with much research covering case-based methods. This method has been expanded on significantly by much research, proving that it benefits from offering explanations relating to the recommendation (McCarthy *et al.* (2004b), explanations stemming from early work in CF by Herlocker *et al.* (2000)), allowing multiple critiques in a single stage of the dialogue (Smyth *et al.* (2004)) or dynamically generating such possible critiques (McCarthy *et al.* (2004a)). Critiquing and other feedback mechanisms can lead to a lengthy dialogue process, which McGinty and Smyth (2002b) examined and found ways to reduce, in order to reduce the user effort required to recommend.

Research has produced a wealth of examples of conversational applications (Averjanova *et al.* (2008); O'Donovan *et al.* (2008); Alon *et al.* (2009)), and approaching item suggestion as a process of conversation has been notably beneficial in case-based reasoning systems (Göker and Thompson (2000)). McNee *et al.* (2006) discussed Human Recommender Interaction, studying ways in which recommenders can facilitate better interaction and ultimately better recommendations. They investigated why users come to recommender systems and through their investigation identified three 'pillars' of successful recommenders.

Dialogue This is "the act of giving information and receiving one recommendation list from the recommender". Conversational recommenders obviously focus greatly on improving the factors involved in this area. Factors such as usefulness, correctness, usability, salience and serendipity all contribute to providing users with a recommendation experience such that they will trust the system in an on-going manner. Of course new paradigms in dialogue call for new methods of evaluation, as both McNee *et al.* (2006) and Wärnestål point out, citing these factors and user evaluation.

Personality This represents the guise the user projects onto the recommender after a period of time, the feeling they have toward it. This is affected by factors such as how much the user feels the recommendations are personalised for them, how current the recommendation is and how much it will adapt to their feedback, as obviously if a user thinks their feedback will not be accounted for they will not provide it. Further recommenders must consider what sort of risks they take in the number of recommendations they show and how much they are seen to pigeon-hole users into groups that will affect their recommendations.

Information Seeking Task The reason the user is using the recommender. How well the user understood the task, how appropriate and important they think a recommender is in solving it and their expectations will all feed into their satisfaction with the experience.

We defer to McNee *et al.* (2006) for more detailed discussion of these factors. It is worth noting that these factors are somewhat mirrored in the findings by Heath (2008), that expertise, experience and affinity were the primary motivators in choosing a recommender.

It has been shown that users are willing to interact more with recommenders in order to participate in a process that is more transparent and therefore fosters more trust in the results (Sinha *et al.* (2001)). In general, there is a tension between making good recommendations and eliciting useful information from the user, as explored by Connaway *et al.* (2011a). It is interesting to note that this is not seen in work such as Viappiani and Boutilier (2011), indicating that conversational recommenders are convenient enough for use. Knijnenburg *et al.* (2011) showed that, across multiple different conversational methodologies novices dislike conversation and domain experts prefer more conversational features. This raises the interesting question of whether there exists a conversational approach suited to everyone including novices. Conversation research so far has taken advantage of the fact that conversational techniques so far have relied on content-based recommendation in order to use metadata to form a conversation. This makes conversational recommendation subject to the same problems as content-based recommendation. Diversity has also been shown to be an issue in conversational recommenders (McGinty and Smyth (2003)). It also remains to be seen if conversational recommenders offer a better experience for any of the specific tasks outlined by Herlocker *et al.* (2004), including the "Just Browsing" task we examine here. Recommendation is increasingly being thought of as a conversation between system and user (Tunkelang (2011)), so this is an exciting area of study.

2.3 Social Recommendation

It has been noted that "the age of the crowd has passed, as public life supposedly becomes ever more virtual, that is to say, organised less around a mass extension of the public square, more around the distributed management of difference." (Mazzarella (2010)). Contemporary social attitudes have shifted to consider mobs and crowds, such as those used in recommendation, as outdated and volatile, with multitudes of unique individuals more in-line with modern considerations. This has much to do with crowds no longer being physical groups of people due to emergent social networking technologies (such as Twitter[10] and Facebook[11]) in the post-Web 2.0 world. The Internet has become a place to share opinions, contribute content and voice uniqueness in real time for its users. This is reflected even in the early research, Terveen *et al.* (1997) for example details an approach to sharing recommendations socially. Social phenomena continue to be examined to account for differences and social connections within technologies that use the power of crowds to offer recommendations, as well as to form accurate information seeking models. The move toward a more social web is having profound effects on traditional information retrieval, as work by Allen (2003) demonstrated, showing that information seeking is heavily influenced by social networks. Terveen and Hill (2001) discussed social recommenders, saying a user "gets access to the opinions of many different individuals; this is a sort of community pulse. Thus, he or she might come across new ideas and information.".

[10] http://www.twitter.com
[11] http://www.facebook.com

Recent proof-of-concept research into social network recommender systems (SNRS) has shown the utility of employing a user's social context when recommending items to them (and are discussed by He and Chu (2010b)). Distinct from social recommenders, which recommend content that exists within social networks (Guy and Carmel (2011)), SNRS use social networks to enforce recommendation through scrapping for content or opinions about content. Bank and Franke (2010) for example processed user-generated content from social networks as rating data. This was done because "reviews are neither objective nor do they represent real quality values". Groh and Ehmig (2007) showed that friends had clearly more similar tastes than independent sets of users, and Ma *et al.* (2011) used social tags to help alleviate issues of data sparsity in recommendation.

There are many ways to exploit social networks in IR, Heath (2008) outlined ways social networks could be harnessed for information seeking, while Morris *et al.* (2010) and Evans *et al.* (2010) both compared the performance of a search engine against asking on a social network, to fulfil information needs, finding that the social networks of many people could sufficiently answer their queries. With respect to recommenders much work has been done, notably discussed by He and Chu (2010b) with social networks. There is a feeling among some traditional recommendation researchers that social recommendation loses the power of crowd-sourcing, or "No matter who you are, someone you don't know has found the coolest stuff."[12]. However, research by Swearingen and Sinha (2001) showed that in their experiments users' friends consistently provided a higher percentage of "good items" and "useful recommendations" compared to (none-the-less still useful) traditional recommender systems. Work by Liu and Lee (2010); Bourke *et al.* (2011) has examined altering collaborative filtering by amplifying friends opinions, which we build upon significantly in this work (Chapter 6). This work hopes to also use socially relevant

[12] Chris Anderson - `http://longtail.typepad.com/the_long_tail/2005/02/why_social_netw.html`

information from sources such as Twitter[13] to explore methods to augment and improve drawbacks inherent in recommendation systems while building more complex social understanding.

One consideration developing from social network research is the issue of trust, involving research questions surrounding whether one user or the system should trust a stranger. This has developed in part from work into making recommenders robust to 'shilling' (Lam and Riedl (2004); O'Mahony *et al.* (2004)), where unscrupulous individuals try to bias recommendations through their actions. Trust metrics have been applied to users so systems can trust their ratings accordingly, and research in social networks is exploring the same concept for users (Golbeck (2005)). Trust has been looked at as a way to combat rating sparsity (Massa and Avesani (2004); Massa and Bhattacharjee (2004)), while Abdul-Rahman and Hailes (2000) and later Ziegler and Lausen (2004) showed that users form connections with people who have similar interests, in context such as movie recommendation (a domain studied by Golbeck and Hendler (2006)). He and Chu (2011) identified trust issues that could impact social recommendation, specifically being misled by friends with unreliable knowledge or shilling attacks from malicious users. Current research suggests trust and ideas of reputation (i.e. work by McNally *et al.* (2010)) are the most complex social standing examinations yet done, but relationships such as influence between different types of people, close friends, independent users or experts, have yet to be examined.

Not all obvious applications of social information are beneficial however, recent research by Muralidharan *et al.* (2012) showed that Google search results annotated with social contacts "go largely unnoticed by users in general due to selective, structured visual parsing behaviours specific to search result pages". While search may not be able to benefit from social annotation, recommenders have made great use of it, for example work by Bogers (2009) to recommend bookmarks other users have

[13] `http://www.twitter.com`

tagged. This tagging is a visible annotation, with the aggregate effort of a group of users to tag content known as a folksonomy, a portmanteau of 'folk' and 'taxonomy'. It is an example of explicit support for social features, rather than implicit accounting for them, such as in views. Social networks have also been used in expert recommendation and identification, for example McDonald and Ackerman (2000a) or in Sha *et al.* (2012) using trends, or Passos *et al.* (2010) using topic-modelling. Work by Amatriain *et al.* (2009) showed that collaborative filtering could be enhanced by use of experts identified in social networks. We briefly look here at a dataset that has experts annotated, and building on such previous work use that fact to examine their part in social interactions, research that could be done on other datasets as supported by these expert-identifying techniques.

While it is true that the common five-star rating system is useful for gauging user interest, it can lead to inconsistencies in profiles as it is intuitively rare for two items of the same rating to share the exact same esteem in the user's mind, a subjective problem that contributes to rating difficulty (Sparling and Sen (2011)). Social network recommenders that recommend from social networks (He and Chu (2010a)) have been a recent hotly-discussed topic and have the potential to elicit useful information not only about social influence, which has been studied by McDonald and Ackerman (2000b), but perceived social influence, which may have a greater impact on users. The hope is that by using social and locational context we can better understand not only what a person wants but also how to ask them for feedback on that suggestion.

Integration of social networking is instinctively useful, since the ways people connect are sure to impact their experiences, and it has been shown by Liu and Lee (2010) that using some social network information improves recommender performance. Additionally entirely new ways of recommending using social networks, such as clustering around social information (Pham *et al.* (2011)), tag recommendation using k-Nearest Neighbour (Gemmell *et al.* (2009)) or realtime mining of socially shared opinions (Esparza *et al.* (2012)) are proving fruitful, ensuring the social web remains an active area of discussion (Mobasher *et al.* (2012)). But new questions and challenges arise as the methods of integrating social information grow and the prospect of drawing on work in other fields becomes more desirable. Sociologists and anthropologists have long studied social relationships and the methods by which they manifest themselves, more complex representations than have been studied in efforts to improve recommendation accuracy. Such work as has been done by Bourdieu (1984) who has examined how different classes and roles develop, finding them to largely emerge from the *expression of taste*, information gathered in volume by recommenders. How a person chooses to present their world, their specific tastes, is a way of depicting status and distancing from groups they dislike. This sort of semantic role, along with the impact such expression might have on others, has not been explored in depth.

2.4 Contextual Recommendation

Context is simply any information that tells us more about the user or process they are engaged in, including sensed information such as time or location, and has grown to encapsulate complex semantic interplays between an item and its environment. People have a cultural understanding of context related to how they use language (Goodwin and Duranti (1992)) and a person's context can be said to be anything that affects that person's decisions, as Yoon and Simonson (2008) have shown. It is also worth noting that social concepts of public and private, such as those related to the sharing seen in recommender systems, have always been intimately tied to representations of context such as location (discussed, for example, by Warner (2005)). Locations define the sort of interactions that are appropriate in them by virtue of how public or private they are, from the privacy of one's home to the public space of a large shopping centre. Savolainen (2010) shows it is a key factor in "everyday life information seeking".

Early work in context for computing was done by Schilit *et al.* (1994). While the use of context has been debated since before sensors to detect it were widely available (Newman and Newman (1997); Lieberman and Selker (2000)) these contexts provide a vital source of information in determining state for social purposes and may well be as useful to a conversational recommender. This has lead to much study (Dey (2001)). For this reason such context is an integral part of social context in recommendation and is examined in this work. Work has been done (by Ingwersen and Järvelin (2005)) to define the variety of contexts that exist, separate from the means of collection. This work was refined and here we present an overview adapted from Ingwersen (1996). Sensed information, inferred from physical sensors and used in such as applications as location-based recommendation, provides a means of detecting some of these contexts.

Intra-object context This context relates to the relationship an object has with itself. It can involve metadata and the connections between item attributes, or the quantifiable structure of the item, particularly of textual content.

Inter-object context This encompasses all the factors involved with relations between items, assigned index terms or external metadata that relates to the item. Playlists are a good example of this, as they connect items in a context that they would not have on their own.

Session context Session context is the context gathered from a single usage or session, a person's usage patterns in the recommender, which involves real user tests or interaction simulation.

Individual contexts This relates to the social, conceptual, emotional or systematic contexts specific to the user. Their impact can be seen in rating behaviour and usage.

Collective contexts This relates to the social, conceptual, emotional or systematic contexts the user inherited from the collective, be it through membership of a community or though being grouped with like-mined users. Though recommendation frequently involves grouping users contextual recommender research has not focused on varying context usage based on these groups. In our work we will touch on this in two ways, in social recommendation the context of other users expressing opinions (Chapter 6) and examining whether one set of contextual factors can be chosen to suit all users (Chapter 7).

Techno-economic and societal contexts These, somewhat more global contextual factors affect all previous contexts, but in ways that can be difficult to detect.

Historical contexts Historical context, the previous events that could influence a person's decision-making.

Currently contextual recommendation work has identified three ways to integrate context into recommendation (Adomavicius and Tuzhilin (2011b)), filtering items before (pre-filtering) or after (post-filtering) recommendation occurs and altering the representation of user-item ratings to be user-item-context ratings. Each has been considered to have advantages and disadvantages, while comparisons by Panniello *et al.* (2009) have shown that neither pre nor post is significantly better, resulting in designs for context in recommenders that are usually decided at build time, with little study of how context are actually used by individuals for the application.

Work by Parra and Amatriain (2011) has shown context indicators can be used as implicit feedback on items with good results. These representational indicators of context are not always enough to truly define the current situation, as recent work by Anand and Mobasher (2007) and Dourish (2004), gives credence to the idea that context is built from a mutual understanding of the current situation through interaction building on such indicators. This, combined with work that highlights that not all context impacts recommendation, Madani and DeCoste (2005); Baltrunas *et al.* (2011) shows that it is desirable to research how to choose the factors that matter to users and comfortable methods of sharing that context.

Derrida (1976) famously said “There is nothing but the context”, highlighting the importance of understanding surrounding factors in understanding the person. Accounting for context in recommendation is hugely desirable, as we have shown, research suggests that it improves accuracy. Research has already pointed out that context is of value in harnessing the explosion of additional information brought about by the realtime social web (Noulas *et al.* (2012)).

Work by Google reports that 70% of smartphone owners use their device while shopping, and the majority of shoppers use online resources for research and purchase in their local store (Google *et al.* (2010)). Mobile applications have been developed that prove the viability of item suggestion in a mobile context (Brunato and Battiti (2003)), and of using location to inform suggestions (Yang *et al.* (2008); Brunato and Battiti (2002); Park *et al.* (2007)). These factors point to a future of computing in a retail context that will benefit from the personalisation ability and interaction offered by a recommender that is contextually relevant. Research by Schmidt *et al.* (1998) warns against the focus on location as a quick and easy contextual factor while missing out on the multitude of other contexts, both sensed and surveyed, which are possible. Interestingly most contextual recommendation work treats contexts as continuous variables, while work by Anand *et al.* (2007) shows that discreet “finite states” also work, but have not been widely studied. Here we will investigate which method users prefer when expressing context.

It is far from simple to use context, as Dhar *et al.* (2000) showed that even time pressure for example has a huge effect on other contextual features and how they are perceived. As previously mentioned recent research Wilson (1999a) has defined three major methods for incorporating context into recommendation algorithms. These three methods are pre-filtering, post-filtering and altering the user model. The drawbacks of these methods in traditional recommendation is that none provide a method to determine which contextual factors are of primary importance dynamically, which is what we study here. Since CF recommenders work by forming groups based on user information any new information has the potential to further subdivide groups, and since recommendation quality is directly related to the size of these groups context must be intelligently managed. In essence there is a risk of creating a "contextual long tail" (by stating that users only share an interest if that interest is rated highly in the exact same context), which has gone without study.

Ingwersen and Järvelin (2005) argue for a breaking down of the division between quantitatively-oriented IR and qualitatively-oriented IS in the study of context due to its complexity and dependance on user sentiment. We explore in this work both qualitatively and quantitatively, the attitudes users have toward context and the use of multiple context factors in recommendation.

2.5 Summary

Currently there exists no coherent attempt to examine social information in the context of the fledgling field of conversational recommenders, or to study how this affects a user's approach to them. Conversational recommender research to this point has not investigated ways to make conversation possible in the most used recommendation methodology, collaborative filtering, begs the question of whether the approach is usable in such environments. Along with this the suitability of conversational recommenders for recommendation tasks such as browsing is not well studied, nor has user response to recommended items from CR systems been studied. The social web and contextual recommender research fields show strong potential to produce a wealth of data that could be integrated with these recommender systems, but as yet there have been no attempts to study these sources for possible complex social or contextual relationships. We will investigate these sources, looking at social context for its ability to predict influence and expert sway, and looking at context for the ways in which users see benefits from different sources. In this way we hope to better understand the impact these sources have on recommendation and lay groundwork for how they might be used in conversation, if indeed they are suitable, in the future.

In summary the major goals and motivations of this research were to:

1. Enable recommendation that engages and responds to conversation without being limited by metadata.

2. Find ways to study how people make use of conversational recommendation to fulfil information needs.

3. Use social context and interaction to better understand what users need and how they view the opinions of others.

4. Build a framework within which user benefit from context can be seen at a per-user rather than per-task level, in order to show context affecting context.

Chapter 3

Conversational Recommendation

3.1 Introduction

Of primary interest to us in our work investigating engagement and understanding is the method of that engagement. We see from multiple sources (Ricci *et al.* (2011); Resnick and Varian (1997)) that computer-driven recommendation works by building an understanding of the people it recommends to, and pairing that with knowledge of the item catalogue it has access to. As we have said in Chapter 2 recommendation algorithms can be viewed as functioning like a bookseller in helping people find things they might want but don't know. To actually engage with the customer, however, a bookseller can ask all sorts of questions that might inform their recommendation, taking an active approach rather than a passive one to gathering data. This process of conversation is the area which we develop in this work.

Our work is specifically interested in investigating a method of conversation that has largely gone unexplored. It is known that conversation offers multiple ways to gain a better understanding of people's preferences as show by Shimazu (2001), but this has been considered a challenge in situations where the person has less domain-specific knowledge (Knijnenburg *et al.* (2011)), whether experience or training. Equally challenging is designing a computer system, both algorithm and interface, that can afford discussion about items it has no detailed knowledge of, as is the case in the commonly used CF approach to recommendation. Since the process of discovering items in contexts such as shopping is an instinctive one the knowledge-focused approaches taken so far not only make it difficult for less *informed* people but potentially fail to account for important emotional reactions to items that could be captured and used for a better browsing experience.

Collaborative filtering (CF) offers the benefit of not needing intrinsic knowledge of items, but this makes conversation difficult. How do we engage people when there is no topic, without offering a very poor conversational experience? As we have mentioned earlier, attempts so far to explore conversation in CF (Rafter and Smyth (2005)) have fallen back on item metadata for the conversation, essentially losing the power of CF to recommend items about which little is known, by requiring that information be present. We explore systems built to offer interactions that require no burden of knowledge on either the user or the system itself. These systems show improved accuracy and the potential to be combined with the many deployments of collaborative filtering currently in existence without the need for augmenting the item catalogue with metadata. There currently exists no exploration of conversation that makes use of only the traditional CF understanding of items through user ratings alone.

The outline of this chapter is as follows: We first look at an approach which in the spirit of CF uses peoples' ratings to talk about items in section 3.2. We have applied this approach to the MovieLens dataset and verified its performance through experimental analysis (sections 3.4.1 and 3.4.2). We further our examination of how conversational recommendation can be extended using information that is not intrinsic to the items being recommended in section 4. This idea of extending the recommendation with extrinsic data will come under further scrutiny in our later chapters. We looked at a conversational methodology that recommends combined sets of items in a highly interactive environment to allow people to browse data.

Conversation revolves around refinement based on personal values and extrinsic metadata. As an application of this approach we implemented a recommender for jogging routes, a task that made use of the routes of experienced runners shared on MapMyRun[14]. Jogging routes are sets of points that provide a run that is pleasing to the user, a quality fulfilled by information beyond what is captured by the points, for example run difficulty, sights seen or goals met. We designed an interactive system that allows users to explore the space, provide pre-conditions for their recommendation and alter the recommendation afterwards to provide feedback, to explore sensible ways to recommend sets of items.

3.2 Collaborative Conversational Recommendation

Recent work by Tunkelang (2011) has shown the value of treating recommendation as a conversation between user and system, which conversational recommenders have achieved by allowing feedback like "not as expensive as this" on recommendations. Our research focuses on creating a viable conversational methodology for collaborative filtering recommendation. Since CF algorithms do not have an intrinsic understanding of the items they suggest they have no obvious mechanism for conversation. Here we develop a means by which a recommender driven purely by CF can sustain a conversation with a user. In our evaluation we show that it enables finding items that the user wants, more effectively, and without requiring any specific training or knowledge of the area.

Recommendation involves finding items that users might like based on what is known of their interests. As suggested by Herlocker *et al.* (2004), six different uses of recommender systems exist:

Annotation in Context This task focuses on providing additional information in context that a user might need. An example might be a system to recommend

[14] `http://mapmyrun.com`

citations as a writer writes an article or paper. This task is not the focus of our work here.

Find Good Items This is the task most often associated with recommendation, finding a list of items the subject will like, often without explicitly stating the predicted score or degree of *likeness*, and with a focus on high quality over complete recall. These best guesses are important in conversational recommendation as it appears here.

Find All Good Items Some tasks have a clear need for a complete set of recommended items, such as legal cases or patent search. In these cases finding desirable items is combined with a priority for recall in order to catch all the items that should be recommended. This is not investigated in our work.

Recommend Sequence This has been described as recommending items in a correct sequence over time, or in a certain order, such as a reading list to familiarise someone with a topic, or a set of items such as a music playlist. We explore a novel method of set recommendation later in this chapter (Section 4)

Just Browsing This is the act of browsing data that allows for serendipitous discovery of items. In these browsing scenarios, the quality of the user interface is usually deemed more important than recommendation accuracy. We are interested in the human-recommender interaction that occurs in order to offer quality recommendations in terms of both recommendation accuracy and user interface.

Find Credible Recommender This task is defined as the user's testing of a recommender system to show that it provides satisfactory recommendations that seem to be credible, and it is the credibility as much as the recommendation that is important. This task is not applicable to our work.

We wish to explore the serendipitous item-discovery of the “Just Browsing” task in a conversational context, to explore the possibility of capturing the casual shopping experience in the real world, where impulse buying is common and finding good items is as opportunistic as it is unpredictable. Since conversational recommendation is useful for differentiating between a person’s immediate and continuing interests we examined employing conversation in widely used CF recommendation.

One of the biggest challenges in recommendation is capturing a person’s unique characteristics in order to model them better and to give better recommendations. It can be difficult to determine if recommendations are optimal where the user can only indicate a degree of success tangentially, which they do by sharing their rating of an item they already have experience of with others. This means people receive recommendations in a session, then have to either leave the session to experience them before being able to give feedback or ignore them. It is hoped that after collecting a sufficient number of such ratings the system can begin to offer reasonably accurate recommendations. This is not the only possible method however, as it has been shown that users are willing to interact more with recommenders and to participate in a process if it is more transparent, which fosters more trust in the results as shown by Sinha *et al.* (2001). Such interactivity can be hugely beneficial, so the question that drove us was how can we best capture these characteristics in order to embody both their interests and their current context.

Information seeking is aided or hindered directly by the affordance of the interface the user interacts with, affecting how feedback can be expressed. A user’s current needs will determine their entire approach to a system, and while much work has been done by those in *information science* to model such interactions in search for example by Marchionini (1995) and by Järvelin (2011), the problem of creating suitable functionality and the interfaces to support that functionality in recommendation, continues. The usual recommender systems interfaces will list predictions of items which users may be interested in (Resnick and Varian (1997)), and this offers little encouragement to elicit user feedback.

Another factor we considered in folding the recommendation process into information seeking was that for any given list a user can only be expected to rate the items they have detailed experience of, with no case for feedback on unknown items. In addition, a recommendation list can be ambiguous as it is not clear what can be done with it to positively influence the recommendation or even to exert agency within the process. Because of this, while recommendation is a ubiquitous part of the online shopping experience it is most frequently seen as an accessory function; users are familiar with the "customers who bought this also bought" panel as the primary manifestation of recommender systems. Ratings and reviews, which play a key part in recommendation are frequently seen as "sharing opinions with other users" rather than "helping the system learn about you". Amazon have attempted to alter this with their "Betterizer"[15], which gives the option to "like" items so the system understands you, but it makes no attempt to teach the user that ratings are the method by which customer suggestions learn about them. Researchers have provided recent re-imaginings of dedicated recommendation systems to better allow people to browse shop items of interest to them, including "conversational" systems that engage users in order to encourage feedback, using methods like asking or proposing items (Shimazu (2001)).

Item suggestions remain an automated background task that contributes additional information to an otherwise directed task. Recent research by Averjanova *et al.* (2008) has taken to exploring methods by which recommendation could be the focus of a system for allowing users to more freely exercise their will based on preferences. Methods like critiquing items based on their properties (McCarthy *et al.* (2004b)) and interactive recommendation (McGinty and Smyth (2002b)) have formed the basis for "conversational" approaches which allow for exploration and an active approach to recommendation thus reducing the pressure on eliciting infor-

[15] `http://www.amazon.com/gp/betterizer`

mation by making it a primary focus.

These methods of critiquing and interacting are useful in establishing that computer-driven recommendation, with its background in predicting a user's interest *a priori*, can benefit from the direct interaction that happens when people suggest things to each other. Conceptually, if users have ways with which to engage with the system which are more than just sharing opinions on what has been seen, we have the opportunity to learn more about them. This flexibility results in a much shorter time to produce accurate recommendations (McGinty and Smyth (2002b)) and diversity in results (McGinty and Smyth (2003)), however it is of limited use to people with a lack of knowledge of the area.

In the work we report here, we explore a new approach to conversation within recommendation. We have developed a way to generate conversation around a large dataset, allowing users to navigate their recommendations in situations where metadata about items is not present. An application called *MovieQuiz*, which allows users to quickly browse recommendations to refine the initial recommendation given to them, is used as a basis for an evaluation of our approach. The approach we take makes use of no special metadata associated with items and as such we felt it appropriate to use the *MovieLens*[16] dataset, a collection of movies built for the purpose of recommendation benchmarking such as this. We recorded user interactions, ratings and responses to a follow-up survey for the purposes of evaluation and we show the ways in which our interactions improve a user's ability to browse the collection and find good recommendations.

[16] `www.grouplens.org/node/73`

3.3 Design and System Outline

3.3.1 The Interaction Approach

Our approach centres around the idea of users choosing their area of interest. We hypothesise that using only the number of ratings and the average rating of items we can reduce the set of items to recommend from in order to offer better recommendations. We provide a means of giving feedback based on the response, either reasoned or reactionary, of "I'd prefer this to that". While this reasoning is fuzzy, imprecise, and difficult to capture it is nonetheless an important part of decision-making for users. In contrast to early work on case-based conversation by McGinty and Smyth (2002b) this is not the same a expressing "I'm interested in more like this", rather the process unfolds like a conversation in which indicating a preference produces potentially entirely new recommendations. Our approach also differentiates a person's *immediate* interests, i.e. in this particular interactive session's preference indications, from their *continuing* or on-going interests, collected when they rate items. It does this by modelling a user's continuing interests using ratings as is traditional but in an immediate session pairing this knowledge with only a subset of the catalogue that they have designated as interesting. This has the effect of allowing a user to affect change quickly and easily based on immediate interest. Further this iterative whittling of the collection continues to make use of the same underlaying algorithm, therefore avoiding becoming a "top popular" approach.

The strength of CF recommendation lies in using rating information to understand users in comparison to others, to place them in a neighbourhood of peers or find items similar to the ones they like. Our approach uses this understanding of items through ratings, by focusing on how popular an item is, and how well it is rated. If an item is defined as i, the popularity of an item ($Pop(i)$) for our purpose is its rating coverage, i.e. the number of people who have rated it, while the measure of how well rated it is ($Rated(i)$) comes from the average rating:

$$Pop(i) = Numberofratings(i)$$

$$Rated(i) = Numberofratings(i)dividedbyNumberofpeoplewhorated(i)$$

From this, any item in the collection can be represented on a graph of popularity against average rating. This graph is a representation of the collection that is equally valid in all areas with respect to user tastes. That is to say that nothing on the graph can be assumed to be worthless, as aficionados of items such as books or film can understand there are audiences for both well-rated niche items and items that everyone has seen but wouldn't be their favourite.

Our approach works iteratively. A session begins with the user having access to the entire collection of items. Two movies are randomly picked from different areas of the collection, one to represent popular items and another to represent highly rated ones. The *popular* indicative movie is chosen from the movies with at least half the average number of ratings, while the *highly rated* one is chosen from movies with at least half the average rating of the collection. This collection is the items considered to be *of interest* to the user, the set that they are working to decrease at each iteration, starting with all items available. The two options are shown to the user to ask "Which do you prefer?". Additionally, a list of recommendations from the collection is generated and the top five are shown below the question, both to give users a sense that their interaction is having a meaningful effect and to show them new suggestions which they may be interested in. If it is the case that a person cannot state a preference for either movie they can refresh the webpage to see new set of choices in the same popular/highly-rated domains.

Once the user chooses either option we have a relative preference (RP), a statement of "I prefer X to Y", and the set of items from which recommendations and quiz interfaces options are generated is partitioned. This means that at each iteration the

Algorithm 1 Collaborative-filtering Conversational Recommendation Algorithm.

$PopularityWeight \leftarrow 0.1$
$AvgRatingWeight \leftarrow 0.2$
$PopularityBound \leftarrow zero$
$AvgRatingBound \leftarrow zero$
while $SessionNotFinished$ **do**
    for $iteminItemCollection$ **do**
        if $itempopularity \leq PopularityBound$ **then**
            $RemoveitemfromItemCollection$
        end if
        if $itemaveragerating \leq AvgRatingBound$ **then**
            $RemoveitemfromItemCollection$
        end if
    end for
    $PopChoice \leftarrow ItemCollection.popularChoice$
    $AvgRatedChoice \leftarrow ItemCollection.wellRatedChoice$
    if $UserChoosesPopularFilm$ **then**
        $PopularityBound = PopularityBound + PopularityWeight$
    else
        $AvgRatingBound = AvgRatingBound + AvgRatingWeight$
    end if
end while

user is being given recommendations from a smaller pool of items, using the same algorithm We use bounding here, which has been explored in search tasks (Smeaton and van Rijsbergen (1981)) but not in recommendation, especially as a means by which conversation can occur. Here we use lower rather than upper bounds, to signify *least acceptable value*. During the iterative process the user is partitioning the movie collection by the least acceptable number of ratings and least acceptable average score, effectively finding the lower bounds of popularity and quality acceptable to them in their current context.

A new pair of options, with list of recommendations, is posed to the user. The degree to which the items are partitioned depends on the density of the collection and our aim is to reduce the set to produce visible change in recommendations through every action. This continues until the user stops answering questions or until there are less than ten items to choose from, at which point all ten are presented.

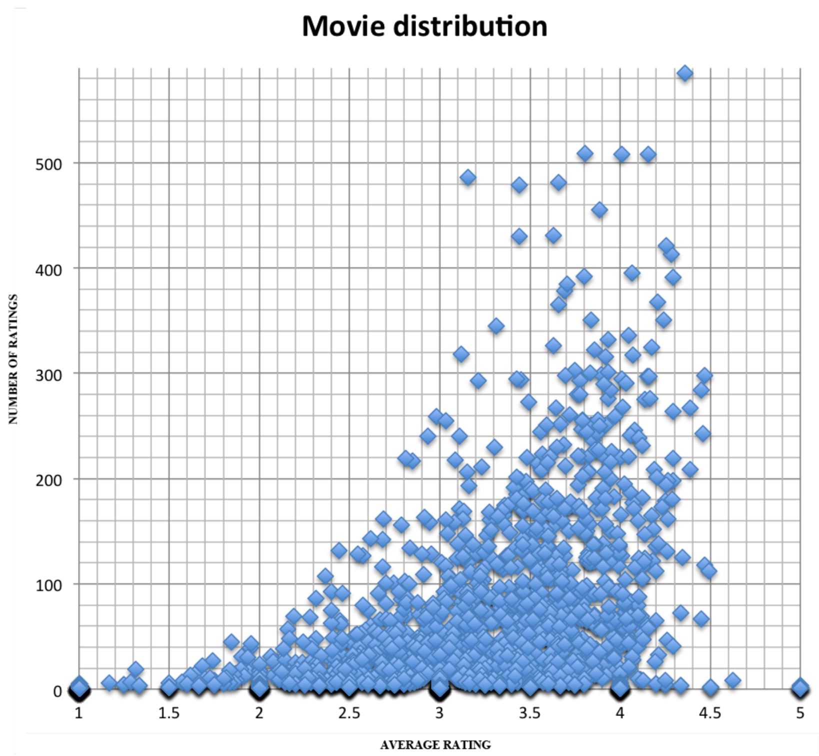

Figure 3.1: The distribution of items in the MovieLens dataset when plotted using our measurements.

3.3.2 The MovieQuiz Application

We developed an application to evaluate our method using the MovieLens 100K dataset which comprises 100,000 ratings from 1,000 users on 1,700 movies. We use this as the seeding data for recommendations, with actual user interaction and rating data collected from other live users. Our example application uses movies which range from "blockbuster" films seen by high numbers of people and "indie hits" that have a very high average score. These two axes represent traits, number of ratings and average rating, that can be judged to be valued in different proportions for different people. Prior to engaging with the conversational interface users were asked to rate 10 films in the collection, from a list presented to them.

The user refines the recommendation provided for them by culling from the collection, movie items which they feel are of no interest to them. The system asks "which of the following two items do you think you would prefer?", to which the user provides a preference which can be used to narrow their possible recommendations. In order to do this without intrinsic knowledge of the items themselves, as CF sees items, we have explored using the information provided by ratings. We guide the user through a series of decisions that subdivide the possible recommendation space according to their relative preferences using a pair of lower bounds, reducing the portion of the collection we dub *of-interest* to the user. This differs from critiquing, where the conversation is based on domain-specific traits. Our approach therefore works with a collection of items that do not have descriptive metadata, making it useful in situations where none exists.

We used a k-NN item-based collaborative filtering algorithm to form recommendations. This algorithm is used for traditional recommendation and we adapt it here for our conversational approach, as detailed above, to recommend from a subset. The adaptation is conceptually straightforward in that we modify it to recommend only films with an average rating greater than or equal to X and with Y ratings, where X and Y are determined by the user's interactions with the conversational interface on a per-session basis. Any recommendation algorithm that can be so altered could be used for this approach.

In order to enable traversal of large datasets by a browsing user, the affordance of the interface we developed allowed interaction while informing the user of the current best recommendations. Our basic layout, as shown in Figure 3.3, is to prompt the user with two candidate preferences. Not shown below the choices is a list of the top five recommended films from the collection according to the current partitioning. Users are given the title and genres of the movie, along with a poster and links to searches for the film on IMDB[17] and YouTube[18].

Experimentally, and as can be seen in Figure 3.1, the MovieLens dataset we use in our application shows a skew toward items with higher ratings. This results in

[17] `http://www.imdb.com`
[18] `http://www.youtube.com`

users needing to express a preference for high ratings numerous times at the start of a session before any significant changes are seen to their recommendations. For this reason we place greater weight on an interest in films with high ratings at the beginning of the process, incrementing the high rating bound by 2.5 on the first action and 0.5 after that for this dataset. The popularity bound was incremented by 150 ratings per action, selecting popular over high-rated items.

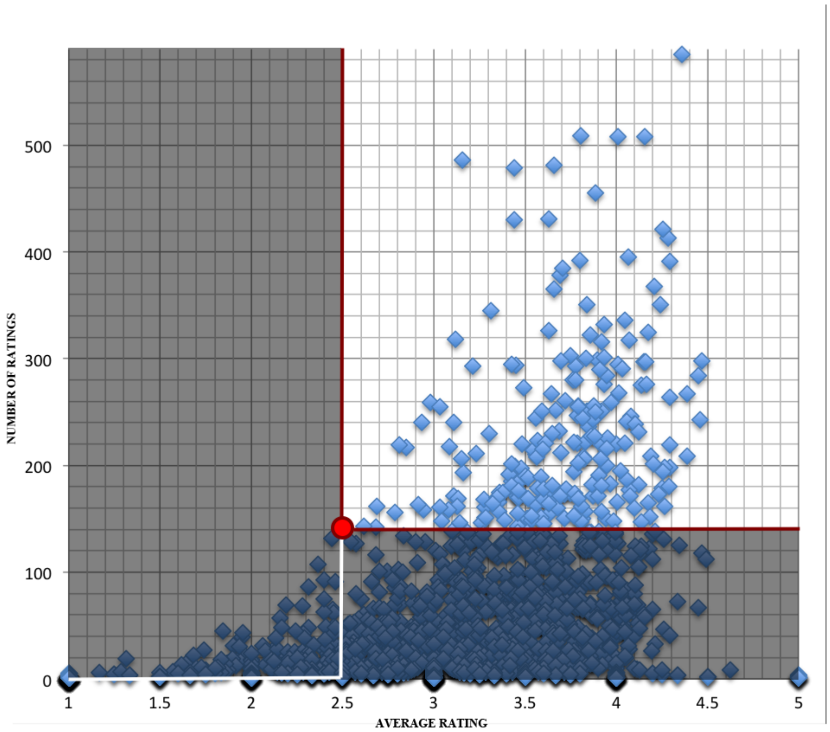

Figure 3.2: MovieLens collection dissected according to the user's choices.

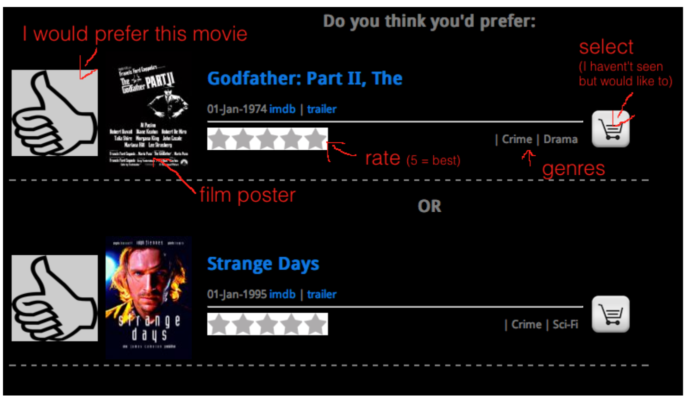

Figure 3.3: The MovieQuiz application interface.

3.4 Evaluation

Our primary interests in evaluation were investigating if our approach helped users to find items they like and if it provides another source of useful explicit context data on a user's interests. To this end we examined interaction logs from an online evaluation using MovieQuiz, and we carried out a user survey to explore how users relate to and make use of the interface we provide. This evaluation made use of the popular crowdsourcing website CrowdFlower[19] to recruit users. These users were required sign-up to MovieQuiz and rate at least ten films. After successfully rating at least ten films the user were accepted as having completed the task. Random examination of the collected data indicates that users explored and rated honestly. Initial unused tests also showed that front-loading the effort, i.e. telling them beforehand they would have to rate a non-trivial amount of items, users were less likely to attempt to shill the system for their own profit. A further follow-up survey, the results of which are discussed in Section 5.2, was later sent to participants.

Since this conversational recommender does not use metadata, the same metrics that have been used in content-based or case-based conversational recommenders McGinty and Smyth (2002b) do not apply here, as we have nothing analogous to a "query" to gauge query difficulty. The purpose of interaction within our recommendation approach is two-fold: to offer users a method of browsing options more efficiently than static recommendation and to elicit feedback that aids in understanding users. We now describe the user interaction in our system.

3.4.1 Interaction Analysis

We generated a detailed log for each user to help understand their actions within the system, and to explore the effectiveness of our approach. Since the interface shows two options in the quiz (more rated or higher average rating) above a recommenda-

[19] `http://wwww.crowdflower.com`

tion list of five items, we recorded a complete list of options and recommendations. For any given rating we examined where the user would see that item on a static list of recommendations, to determine if interaction helped the user find the item more easily and what the average prediction error of ratings was, i.e. the degree to which interaction corrected the algorithm's prediction of that user's rating of that item. We also considered the average number of *moves* or interactions that were needed to get to an item that a user rated, a measure of user effort and system efficiency not dissimilar to query difficulty. For this set of tests we used the item-based kNN CF algorithm that was in place in the MovieQuiz application, using Pearson correlation to determine item similarity.

We gathered 4,153 intra-conversation ratings from 251 people, and recorded the details of their 2,415 moves within the system. The average number of sessions (complete sets of interactions from start to end) each user had was two, with 9.6 average moves per user. The average user rated 20 items over the course of their sessions, having initially rated ten items from a non-interactive list before starting (which were excluded from our analysis).

Our set of tests involved an examination of where the items that users rated would appear on a flat list of recommendations. In order to test this for each user we used the same item-based collaborative filtering algorithm used in the MovieQuiz application and generated a list of 100 recommendations for them given their initial 10 ratings, made prior to using the interactive interface. Of the 4,153 ratings given while interacting with the system, the recommender algorithm alone lacked sufficient information to recommend 3,704 of the items within the users' top 100. These ratings were therefore excluded from the mean and standard deviation figures generated in Table 3.1. We also generated figures for the number of moves taken to get to an item worth rating, the average rating, and the error of predicted rating given by the algorithm.

Our findings, presented in Table 3.1, show a number of things. If the algorithm

Table 3.1: CF Conversational Recommender Interaction Analysis.

Data	Mean	Std. Dev.
List place	77.9	22.3
Moves-to-rate	2.33	2.26
Rating	3.60	0.41
Prediction Error	3.27	1.15

recommended an item that the user rated, it was in 78th place on the list on average, with a large deviation. This was the case for only 449 items, the rest being below 100th place on the list. Exclusively through interaction our method accounts for a low precision in this way, by subdividing the collection to show items of interest to the user. If the recommendations were listed in groups or pages of ten as search results are, then it would take seven actions (i.e. clicks of "next page") before the user found their item, compared to an average of 2.3 actions in our approach. It follows that our approach would enable users to find the items they were looking for with greater effectiveness.

We then looked at how usefully distinct the ratings were, the thinking being that increased user effort may lead to ratings that are more telling about the user. This would mean that the system can "understand" them quicker, a measurable example of the idea to "empower people to explore large-scale information but demand that people also take responsibility for this control by expending cognitive and physical energy" (Marchionini (2006a)). We calculated the difference between the predicted rating and the actual rating each time a person assigned a star value to an item. We did this sequentially, so the system had as much training data in simulation as it did at the time of rating. We found reasonable accuracy as defined by RMSE (discussed further in Section 3.4.2), though even so the average prediction error for individual rated items was 3.27, a much larger value than the average RMSE, indicating that the items the user chose to rate were either not ones the system would have recommended (predicting the score too low compared to actual ratings) or were recommended when they should not have been (predicting the score too high). These unexpected items could not be accounted for through the algorithm alone, meaning they represented significant valuable information in modelling the user's preferences, and therefore our conversation helped the user find them. The average rating was 3.6 with a standard deviation of 0.4, indicating users expressed opinions on items in a marginally positive way.

Our collaborative filtering conversation helped users find items that were of interest to them in a measurably more efficient way than a static recommendation using the same algorithm. We followed this with an exploration of user attitude toward the conversational approach.

3.4.2 Interactions As A Data Source For Improved Accuracy

In addition to the interaction analysis we performed an analysis of how the relative preferences (RPs, the "I prefer X over Y" ordinal expressions of interest given by users) may be used to improve recommendation accuracy. Since RPs can be collected regardless of algorithm choice we focused on using them in different ways that could easily be integrated into any approach. While the RP data we gathered could be handled like implicit feedback and integrated into a recommendation algorithm in a similar way to Douglas and Jinmook (1998) (especially music, where listens can be more telling than ratings (Celma (2007))) the user is required to explicitly offer a "gut reaction" to the content in a way that we felt might be more valuable if handled in a more explicit fashion. To this end we designed four methods by which the preference data may be used, as a preliminary exploration of this fuzzy explicit data.

During our online testing phase we collected 2,415 RPs from 251 users, an average of 9.6 preference indicators per person. This may be a small amount of data, but this is to be expected as it is gathered from an average of 2 sessions per user.

Table 3.2: RMSE scores of relative preference session data integration.

Training	IB-CF	4/2 IB-CF	3 IB-CF	NN IB-CF
10%	2.65664	**2.51621**	2.57652	2.58181
20%	1.95620	1.93625	**1.92191**	1.93514
30%	1.78158	1.78973	**1.77042**	1.78299
40%	1.76034	1.77471	**1.75826**	1.76015
50%	**1.74209**	1.74681	1.74357	1.74369
60%	1.66055	**1.66046**	1.66684	1.66305
70%	1.51721	1.51873	1.51759	**1.51144**
80%	1.29378	**1.28640**	1.28826	1.28730
90%	0.92503	0.92498	0.92641	**0.91911**

The algorithms compared in the test were the same item-based kNN collaborative filtering algorithm used in the system, and several modified versions. We designed a number of ways in which the relative preference information could be treated as explicit data for easy integration into the system. For each variation we performed five-fold cross-validation to arrive at RMSE scores.

Our first variation, labelled 4/2 IB-CF, is a simple assumption that for each RP the user *mildly* liked the film they chose and *mildly* disliked the one they did not, so we set them as explicitly rated 4 and 2 out of a maximum of 5 respectively. While rating is undoubtedly a personal act, with some users frequently rating 5 and others never rating 5, we wish to see if simply getting more data would prove useful. 3 IB-CF is an approach that marks each chosen item in the relative pair as rated 3, the reasoning being that the user could rate the item rather than "prefer" it, and since they did not they do not have a strong feeling about it, but are aware of it. The fourth approach, NN IB-CF, attempts to re-use the similarity knowledge of collaborative filtering by assuming that the user prefers the chosen item because it is in some way similar to an item they have already rated. Following this we calculate which of the user's already rated items is most similar to the chosen item and we assign the chosen item the same rating, i.e. if it is most similar to a 4-out-of-5 rated item it will be given a score of 4.

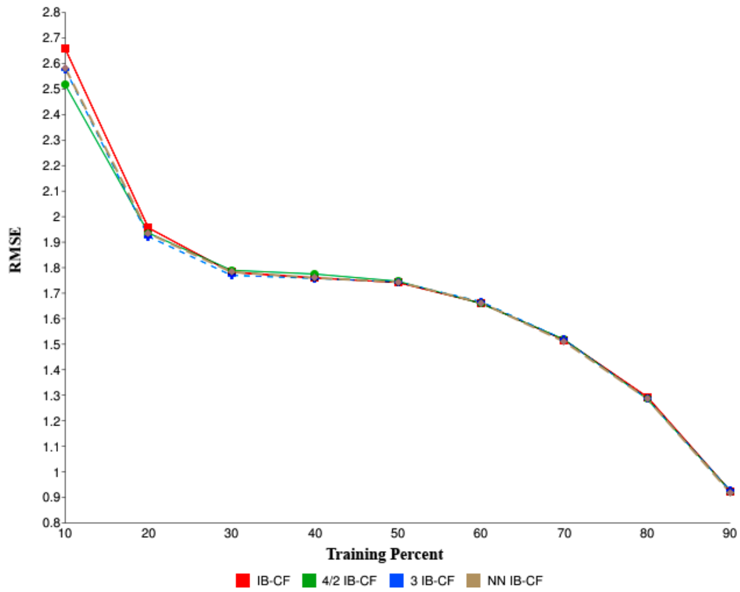

Figure 3.4: RMSE accuracy for the 4 CF CR approaches tested.

This exploration shows (in Figure 3.4 and more clearly in Table 3.2) that even the small amount of extra explicit data made available through a single session with an interactive system improves the accuracy of the recommender, and does so slightly faster. The use of RPs as a method to gain information from new users could help reduce the "cold start" problem surrounding modelling new users which still exists in pure CF systems. It is also possible that RPs could be treated as implicit feedback.

3.4.3 Discussion

We have shown that it is possible to offer conversation in a recommender system using only rating-derived data, a novel contribution that offsets the more usual reliance on metadata attributes for conversation. We have found that users are satisfied with the mechanism we present for responding and finding items without confusion. Also clear is that the explicit information in the form of relative preference statements that can be harvested offers a possible new source of feedback which may be harnessed to gain perspective on user information needs. This approach opens up a new avenue of potential exploration, in that the reasons for choices are not immediately clear to the system, work could be done in "user explanation" to allow the user to explain to the system their interest. Further work could add other, metadata-related dimensions to the refinement to combine it with traditional conversational recommenders, for a "I want to be recommended only whiskeys older than 5 years with lots of very good ratings.". Measures like controversy, the range of ratings an item gets and others related only to rating data could also be explored as future axes along which conversation can occur. We chose "popular" and "well-rated" as axes in order to maximise the usage of available data in our experiments, and since the items were unevenly spread boundaries were adjusted at different rates.

We explored feedback from users of an application designed to prompt interaction, finding users greatly prefer an interactive interface to being given a list and had no trouble making choices to provide feedback and, in their own minds as well as demonstratively, improving their suggestions.

Recent research by Knijnenburg *et al.* (2011) has found that specific domain knowledge correlates with a preference for more interaction in recommendation, but here we have shown that a greater degree of interaction need not come with a domain knowledge barrier, provided it does not hinge on domain-specific attributes. We have shown this technique works to an acceptable level, however since existing conversational recommenders are gauged on their abilities relating to a case-base, i.e. using query difficult, a measure that is impossible in CF, we cannot easily compare the performance of our approach against them directly.

The work of recommendation systems is felt in numerous aspects of popular culture, from Internet shopping sites to Facebook updates. Some have been hesitant to rely on it for fear of the so-called "filter bubble"[20], the idea that they will only be exposed to a narrow selection of things the recommendation algorithm judges

[20] `http://www.ted.com/talks/eli_pariser_beware_online_filter_`
`\discretionary{-}{}{}bubbles.html`

they will agree with and their world-view will therefore be limited. While this is an overstated problem (recommendation never filters but simply ranks and personalised content can still be browsed), the legitimate issues related to a lack of dynamically changing options are important not to ignore. Our work suggests a solution to incorporate interaction to allow a user to sift through false positives to the lowly-ranked range of information they actually want, as is shown in Table 3.1, people are able to find items they want, that would otherwise be ranked lowly on a list, through fewer (2.3 instead of 7) interactions. The work here is designed to work seamlessly with CF, meaning it will generalise to any application of CF, including music, books, or collections of mixed classes of items such as Amazon's shop.

From the user's perspective we have offered an entirely new way to receive recommendations, which allows them to browse a large number of personalised information quickly and transparently. By engaging people in conversation we improve their ability to find items, in an open way. Given that privacy and the use of personal information are growing concerns in the public eye this transparent approach might also improve user satisfaction with how they are modelled in a recommender system, giving them transparent control over the process of modelling. By designing a conversational method for the least content-rich recommendation approach we have created a method that can in future be incorporated into any recommendation algorithm to allow for interaction without domain knowledge.

Importantly our work here pointed to an interesting conclusion, that is, people do not necessarily feel hindered by their own lack of knowledge within a domain if a conversational process is designed not to question them on that knowledge. In other work by Knijnenburg *et al.* (2011), the conclusion was that people with less domain knowledge like less conversational systems, but this seems to be caused by focusing the conversation on domain knowledge; asking a user to critique the focal length of a camera is difficult if a person has never used one. By capturing reactions of relative preference people of all levels of domain knowledge can contribute to a conversation that improves the recommendation for them. Further, the recommendations within the system are derived from a collaborative algorithm which does not take metadata about an item into account, often generating serendipitous recommendations. A benefit of this, for example, is a user liking a film such as "Inception" which might lead them to a film, for example "Lord of the Rings", that everyone who likes "Inception" also loves. However from the user's point of view this relationship may be unclear; they will look for feature similarities between the two movies and find very few. This is highly related to the problem of explanation in CF. The question of how this will affect the user's perception of the system, and how to deliver these recommendations in a way that makes sense to the user, is an important issue.

From this series of discoveries we became interested in modes of conversation that offer improved recommendations without requiring domain knowledge. This led us to explore the task of recommending running routes in an unfamiliar area, using a combined case-based recommendation.

3.5 Comparison to Related Work

In creating and testing approaches to conversational recommenders we have contributed to the larger body of recommendation work. Here we discuss this related work with respect to our contribution in order to contextualise it within the current state-of-the-art.

3.5.1 Collaborative Filtering and Conversation

Recommendation is traditionally regarded as an information retrieval problem in one of two broad forms as shown by Ricci *et al.* (2011), collaborative filtering (CF) and content-based (CB) recommendation, as we discussed in Section 2.1.1 and Section 2.1.2. CF recommendation attempts to mimic "word of mouth" suggestions, those recommendations users would expect to hear from their friends, by finding people like themselves whose similar tastes can be used to offer likely good items. Recent research has highlighted the need to treat the recommendation process as conversation, an interaction between the user and a system they should trust (Tunkelang (2011)). In such research, conventional recommendation is paralleled with a conversation, outlining a respectful process that does not place heavy cognitive load on the user by respecting other content it appears with. This shift in approach will highlight that users' rating information provides a better recommendation, rather than being just a mechanism for the user to share opinions with a community. Researchers have looked at implicit feedback, such as items viewed or the time they are viewed (Hu *et al.*), as a way to infer interest without direct user engagement. In interactive or conversational recommendation, as we discuss in Section 2.2, this is taken further, with the aim to "empower people to explore large-scale information but demand that people also take responsibility for this control by expending cognitive and physical energy" (Marchionini (2006a)). By requiring and rewarding effort or "asking rather than guessing", this is seen as a way to capture what the user likes and the system may more effectively aid information seeking.

Work on ways to make a conversation between a user and a system possible has centred around case-based recommendation. Leveraging the well-described items in a case-base interaction of the form "I want something like this but less expensive, or a different colour", called critiquing, has been explored (McCarthy *et al.* (2004b)) with some success, as has preference-based feedback (McGinty and Smyth (2002b)). Recent research with case-based conversational recommenders concludes that users prefer a level of control that mirrors their domain knowledge, i.e. someone who knows nothing about cameras will not know what feedback to provide on lens aperture, as discussed by Knijnenburg *et al.* (2011). There have also been explorations of recommendation as a game by Alon *et al.* (2009) or from a Human Computer Interaction perspective by McNee *et al.* (2006).

Chapter 4

Combined Recommendation in a Conversational Interface

In this chapter we consider the availability of real-world information on exercise, in this case corresponding to jogging routes, how conversational interfaces might involve a user in recommending routes for leisure running in unfamiliar areas. We describe the Exercise Builder, a proof-of-concept application that helps people to plan their running routes by combining case retrieval, interactive adaptation, and multimedia explanation into an integrated, online service.

Recommendation systems help users to make choices in the absence of either detailed experience or knowledge of the choice options (Resnick and Varian (1997)). They attempt to fill our knowledge gap by mimicking the friend who advises on movies, the book critic whose opinions are always spot-on or the magazine that always gives the best reviews of restaurants. At present, recommendation is almost as ubiquitous as search through its widespread uptake by businesses on the Internet and covering all kinds of services and products. These systems are commonplace as a method for highlighting to users new items such as books, movies, websites, hotels or businesses, which will most probably be of interest or of use to them. Automated recommendation seeks to provide users with accurate and useful recommendations of atomic entities such as a complete book or a movie, a complete website, a hotel, etc., all within a specified and narrow domain. The technology underpinning recommendation systems continues to be based mostly on textual metadata for representing the entities while non-textual media such as image and video has limited use in the operation of recommendation, though non-text entities such as movies may be the objects that are ultimately recommended.

In this section we extend the conventional recommendation process in two directions and we examine the effects on system design. Firstly, we focus on the process of recommendation as conversational interaction for users with a knowledge gap. This conversation helps to refine and focus the user's real information preferences, in much the same way that much of our information seeking activity takes place as an interactive search process anyway. We here examine the role of design in recommendation, with respect to interaction, what effect allowing the user to tweak, explore and variously modify the recommendation has on how they use the system and on system functionality. It is by doing this that we seek to account for the unique interests of a user, in the form of tactic data such as what they value, their contextual desires and similar difficult-to-detect factors, while also offering good recommendations.

User values and user contexts are not easily captured by inference alone and we examine designing the usually non-interactive process of recommendation around supporting their agency. This is a novel contribution because it considers human interaction as key to the recommendation process, not merely base data, or accept/reject responses. The task of traditional recommender systems has been to find users or things that are similar to what the system knows about, someone in order to recommend items to them, thereby forming groups of roughly similar people. The effect is that the more that is known about a person the more effectively s/he can be grouped with others, but their unique viewpoint, their surroundings and the values with which they make decisions, is not supported in any way. We examine how a conversational design impacts that system by allowing the user to directly interact with the system and to stamp their own unique characteristics on the process. In addition we engage in this conversational interaction to further support and allow for the second contribution of our work, which is to do with the unit that is recommended.

Conventionally, discreet units such as books, hotels, or electronic goods are the topic of the recommendation process whereas in this work we recommend a route for a runner or jogger in a new way. We recognise that for the purpose of leisure running, traditional traffic-navigation algorithms do not account for the factors that runners and walkers value such as scenic beauty. Building on work done for route composition, our approach is that the route is an aggregation of parts of other routes which in turn have their own recommendations. We thus build up the object that is recommended, the route, out of fragments of other routes combined together into a new entity. This compound recommendation drawn from multiple sources forms a base, and we design around the user, exploring the space within which the recommendation is given. The resulting approach is to design a way to recommend a crowd-sourced compound entity and to provide worthwhile and useful information for the user to alter the provided recommendation if desired. We demonstrate this with a system we have built and we illustrate its usefulness and feedback from users through a qualitative evaluation. The results of this survey will be examined in terms of the opportunities and implications for designing new recommender systems. We show that not only is it beneficial to provide alternatives as a form of explanation, but it offers new users a foothold in what can otherwise be a daunting domain-specific field.

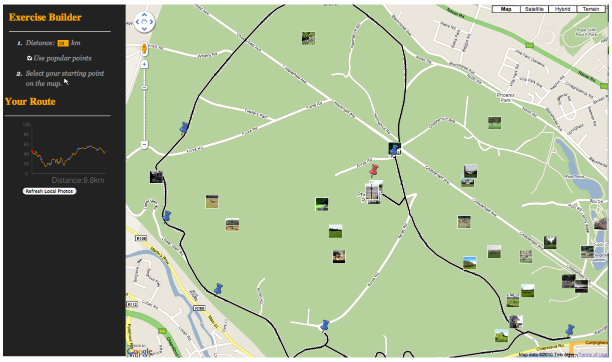

Figure 4.1: Exercise Builder interface, seeking to make route planning for exercise easier.

4.1 Method

Exercise Builder is an application for people who wish to get physical exercise in a new area along routes that experienced exercisers would deem good. The system is in place for such people to plan a run before doing it, either at home or on a mobile device in-field. For this purpose we use Google Maps overlaid with photos of the area to help users of varying levels of familiarity with the area know what to expect and find things that interest them. We also include an informative sidebar and drag and drop markers on the route to make exploring and altering based on desired criteria as frictionless as possible. This is a non-trivial task, as different individual runners might like routes for different reasons, such as beautiful sights seen along the way, particularly challenging uphill and downhill sections or other tacit factors. To mirror this wide variety of motivations among the experienced runners our user, who themselves may not be experienced, may wish to have some influence over the route recommended to them.

We have targeted visitors to a new city or those who are novice joggers unfamiliar with their locality, as the primary user groups for Exercise Builder. This is because these groups have the most need of a service to find routes in an environment they might not be familiar with, whether with regard to specific routes suitable for exercise, or even with the geography of an unknown area. For this non-targeted route planning we propose an interactive model of recommending composite items. This is a model in which users are engaged in the recommendation process from the outset. Users are encouraged to explore and modify aspects of the overall recommendation based on multimedia content presented to them after the initial recommendation has been made. This effectively includes the user in the recommendation process and adds him/her as a real time human data source, able to exert influence based on what he/she values in a good route. The aim is to produce a system that will provide an acceptable recommendation that can be interactively refined based on requirements and preferences that the user discovers, only through exploring the multimedia content which is relevant to different aspects of the overall recommendation.

Such a system as outlined above is designed to interact with users after the initial recommendation occurs, allowing them to weight the current and all future recommendations. It is therefore important that the user understands they are not just being recommended a route for their run/jog, but being lead through a process to build a recommended route based on their preferences. The aim is to provide new runners or walkers with access to the knowledge of experienced runners, which they can use on their run. This crowd-sourcing is done using a case-base of 1,301 routes that were run by running enthusiasts in a given city, and then recorded and uploaded to a popular running website (MapMyRun.com). These runners have expressed in their actions with regard to physically running a route and recording it, that they found it of interest for the purpose of exercise, but there is no associated metadata for perceived difficulty or for interest.

The Exercise Builder is designed as an online application with minimal user interface clutter but a specific aim was to account for the lack of metadata present in the run database by allowing users make judgments on routes. In our system we endeavour to account for a personal expression of interest through embedding multimedia, in this case photos of the area, in the map to allow user judgment to play a role. Additionally we calculate route distance and elevation information to be displayed as data in an informational sidebar. With this information we have built-in a mechanism to give a reason for the user to express their agency, they can find monuments, scenic views or more difficult pathways to suit them. This ultimately allows us to capture their uniqueness and use it in future to recommend trends that others might be interested in. By making recommendation the focus of the system the user is actively tasked with finding the best possible route for them from a recommended baseline, allowing them to establish how they are different from other users.

As mentioned above, the architecture for our route recommendation system depends first and foremost on engaging the user, which represents a shift from the usual application of such recommendation being a feature added to a larger system. In contrast to other systems such as that developed by McGinty and Smyth (2003) or by Göker and Thompson (2000), our system establishes a conversational style by having a linear ask-respond style conversation, thus iteratively reducing the recommendation space. The result is that in a system such as the one outlined below, the user can effectively create new items (routes) that would not otherwise be recommended, which can be saved for future recommendation. It also seeks to allow the user to guide the process more fully using multimedia elements. In this way the user benefits from increased knowledge of the recommendation space and is thus more fully informed as to the quality of the recommendation. This addresses one of the drawbacks of conventional recommender system applications, the issue of how to resolve question in the user's mind of why something is being recommended. Sometimes, feedback along the lines of "Users who bought X also bought Y and Z", just isn't enough.

Since the architecture is designed to focus on post-recommendation refinement, explanation and information solicitation, the pre-recommendation information requirements can be relatively simple, indeed the system can benefit from a certain ‘pacing’ of information gathering, with too much initial form-filling becoming tedious and hindering usage. The ideal format mimics a conversation, with the user providing the system with a relevant piece of information such as, ‘I do prefer running on grass so Central Park (New York) would be good to include’ or ‘I’ve already seen the Coliseum last time I was in Rome’ and the system renewing its recommendation to reflect this.

Recommender systems by their nature will group or stereotype an individual, which makes it quite difficult for such users to be able to express individuality quickly. Conventional systems are designed for applications such as supplementary product suggestion where the goal is a long-term modelling of the user, and the user does not have to confront failures. Here we have worked on an approach using interaction, as it seems an appropriate mechanism to capture here-and-now context as well as core priorities of users.

Context, information about the user’s environment, has been shown to affect choice directly as shown by Dhar *et al.* (2000). In recommendation, context has presented an interesting problem. It is a challenging problem because for different applications, context will matter for different reasons. Body temperature plays no part in movie recommendation but plays a key role in health analysis. For the Exercise Builder we have not employed direct sensory intervention, so we seek to allow context to play its part through interaction. Users are free to change routes based on immediate contextual needs or their less changeable priorities, though we do not distinguish between the two motivations.

The application seeks to tap into the knowledge of a community of runners to provide tacit knowledge about scenic beauty and run difficulty (that has no means of being captured otherwise) without specific knowledge of the area, to show what the community as a whole value for its runs. It then balances this by handing power over to the user to tweak this route to their desired one, whether based on their current context (e.g. halfway through a training regime, need more uphill sections) or values (as one survey participant said “I prefer to run to a landmark as a goal”). Those that value scenic routes can evaluate this aspect of the route through the photos embedded in the map.

A primary concern was domain knowledge, as we wanted to study the utility of this model on groups including those without knowledge of the area or of running in general. Exploration is meaningless if novice users are unassisted, so the technologies we used to build the system support an attempt to make the area more worth exploring. To this end we embed photos of the area in the map to allow them to explore. This metadata covers both the route and any other potential routes in the area.

As a fitness-focused application this seeks to be as tactile as possible in oder to engage and hold a person’s interest in their routine. This ease-of-use is facilitated by support on multiple platforms. We have tested the application on desktop computers through the browser and on mobile devices, specifically the iPad 2 and Google Nexus One Android mobile phone. As far as we are aware this is one of the first health-based recommender applications, with only the work reported by Miyo *et al.* (2007) appearing to study similar areas.

This focus on a variety of devices, touchscreen, or desktop, allows planning in a wide variety of situations to fit with the varying routines of users and enables us to study interaction on various platforms. We designed the Exercise Builder to be used as a precursor to a run, a process that can happen in many different situations for many different users. As such we have built our application to be accessible in many different contexts, to allow it to meet the requirements of planning runners. To do this we tested and developed the interface for desktop use, for planners working at home or some time prior, and mobile use, for use in situ.

4.2 The Recommendation Architecture

Our approach uses case-based recommendation to compose sets of route-points forming good coherent recommendations to users in new cities. It follows the CBR cycle in that it operates in 4 phases.

- In the retrieval phase cases are retrieved that have similar preconditions to the current problem. Here our system collects routes in the locale that fit the user's ability, using their desired distance and start-point as the basis.

- In the reuse phase the system evaluates how appropriate a case is to the user. This is where our system finds points within routes and plots the combined recommendation into a single coherent route. An appropriate case is one which has a point within a kilometre of the user's start point and is within a kilometre of their desired distance. If one is not found a compound recommendation is formed from other routes, as explained in Section 4.2.1

- In the revision phase, i.e. the relevance feedback of the user, the system evaluates the user's interest in the new item. Here we explore the idea of offering extrinsic data, information about the area around the route, not the route itself, to allow the user to understand their recommendation and what might suit them better.

- In the retain phase useful information is saved to improve future recommendations. Here our system saves new routes created through interaction to be recommended in future.

The case-base that we draw on for these recommendations is a set of running routes. These routes have been run and recorded by actual runners, indicating they are viable options for running. Each run has attributes of distance and a list of points associated with it. Each point has a popularity value relative to how often it is actually run. Using this case-base our approach recommends new routes composed of route-points to users. In this section we will describe in detail this process.

4.2.1 Initial Recommendation

The method by which the initial route recommendation is made is to choose a route which is a hybrid of collaborative and content-based recommendation. A set of points is constructed from the user's stated preferred running distance and initial starting point. Routes are comprised of a set of GPS points detailing the route taken as well as metadata, distance and popularity (the sum of each point's number of occurrences in other routes). Routes are similar based on their length and popularity, and recommended to a user based on that user's preferred starting point and allowable distance. The system first finds the set of points constituting the most popular route that is not greater than the user's running distance within a kilometre of their starting point. If this route alone is of insufficient distance (a greater than one kilometre difference) the system appends to this the set of points of the most popular route not greater than the difference. The resulting set of points is an aggregate of one or many routes that is the desired distance for the user. The average route in our sample database contained 92 points, which proved to be too much for users to interact with in a meaningful way, so from this set, eight evenly distributed points are chosen. The route is then built by Google's DirectionService using these points and the start point, and displayed to the user. The end result is a route combining elements of potentially a number of routes and the user's start point.

4.2.2 The Interactive Multimedia Component

Importantly since we are not seeking to optimise for close distance, but for desirable points along a route that approximates the desired distance the application must

Figure 4.2: The red pin designates the route start, blue can be moved to modify the route.

make altering the route easy for both desktop and mobile users. It does this by making use of the eight waypoints along the route, where pins are placed. These pins can be dragged to a new desired waypoint, which recalculates the route to take encompassing all the changed waypoints. This allows for a tactile user experience, as it supports both mouse interaction and touch screens.

After the initial route recommendation is made, the user is shown the touristic and other interest point attractions that lie on, or close to the route. This, along with a graph detailing the elevation of the route, serve as explanatory notes giving the user an idea of what is in the area and why the route is being recommended. The use case here is for a user who is unfamiliar with the neighbourhood of the route, perhaps a visitor to the city, and so s/he may wish to take in some of these landmarks while on the run/jog. For example, while in Beijing we may want a route that takes us past the Bird's Nest Stadium, in Washington DC we might want to cover part of the National Mall area or in London it could be Tower Bridge. The approach taken is to offer the user the chance to browse connections between metadata either intrinsic

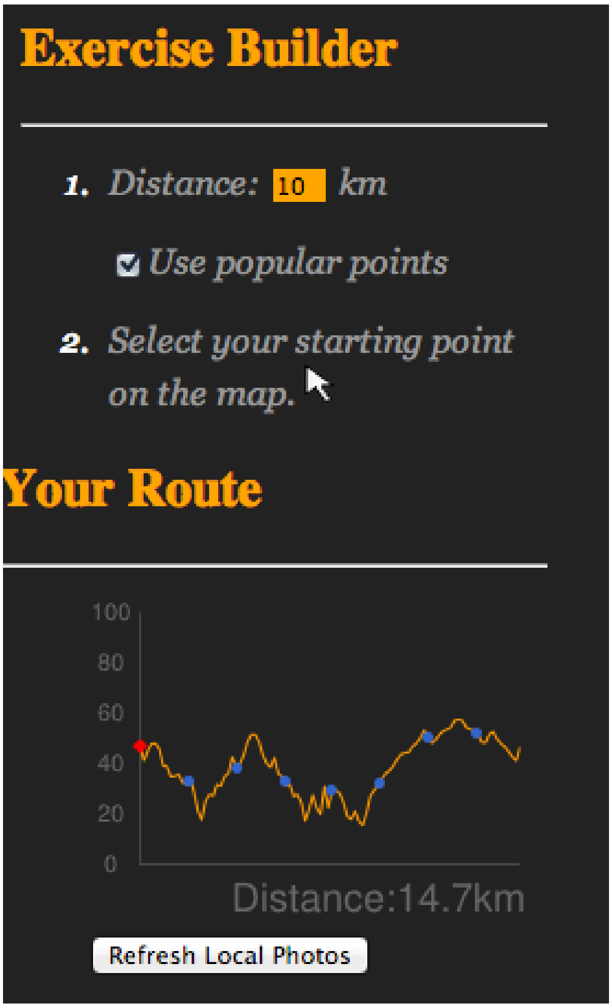

Figure 4.3: Exercise Builder provides information about route difficulty through elevation.

to the item (in this case media on the route) or closely related to a property of the metadata (here near the GPS coordinates of other media). If users frequently modify their route to run close to monuments for example these will become more popular and therefore more recommended. This can be considered a hybrid recommendation technique that prompts the user with recommended items and then allows them to refine that recommendation through their interest in specific metadata (which could be generalised to music genre, screen-time of actors or how frantic the trailer was in other recommendable items).

We use the initial route information to gather a collection of multimedia content, which is then presented to the user. In the Exercise Builder demonstration system

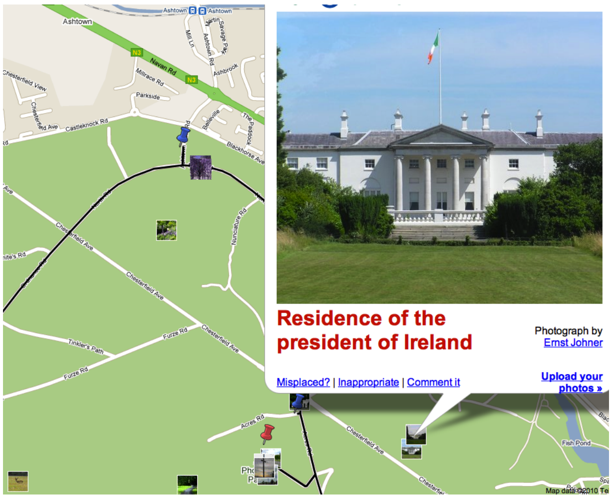

Figure 4.4: Exercise Builder's embedded photos can be interacted with to see larger versions.

this multimedia content comes in the form of a layer of embedded photos. These photos come from Panoramia, a site that provides location-tagged images uploaded by their users. These can range from holiday photos to landscapes, all of which contribute to the user's understanding of the geographic area. The Panoramia site provides an API to select popular images, from which our system takes the top 50 images that were taken within the visible map range, essentially returning a combined set of images describing the recommended route and its currently visible alternatives. This set of images is used to inform the user of both the sights they will see on the run and potential sights that are nearby that they will nonetheless miss.

By allowing users to view multimedia information describing landmarks near their route, these elements become metadata for the human element of the recommendation system to evaluate. The user can reject or make alterations to the suggested route based on information they learn of by exploring content such as photos, trailers, video reviews or related audio. This involves the user in the recommendation process, effectively providing the additional information a user needs in order to make an informed decision on the quality of the recommendation as it relates to them specifically. This functions similarly to explanations in other recommenders, but with the addition of offering explanations of areas for which there may be no route in the case-base. It also enables them to demonstrate their unique interests directly, without having to wait for a user history to be built.

The user is engaged in an interactive exploration of multimedia related to possible route recommendations, allowing him/her to modify the initial recommended route. This is done in a map-based interface with the current route recommendation highlighted, some metadata about the route such as the distance, altitude profile, estimated time to complete, etc., included. The act of changing the route via a drag and drop action on one of the 8 drag-points on the map interface can be regarded as creating a new route, and acting as a form of explicit relevance feedback for recommendations of landmarks to be included, with the benefit of potentially adding to the recommendation corpus.

4.3 Evaluation

The Exercise Builder was used by a group of 66 users interested in exercise, and each was given a complete brief on how to use the Exercise Builder with specific instructions on how to browse the area for pictures and how to modify the recommendation should they wish. Given the low number of routes available (the case-base started with 1,301 routes), the experiment was centred on the most popular running area, the Phoenix Park in Dublin. After they had become accustomed to the application the runners were given a short survey to evaluate how they made use of the route recommendation and how the routes reflected their wants and needs.

4.3.1 User Survey

We conducted a user survey online, with users self-evaluating their experience and knowledge levels. Of the 66 users, 15 lived in Dublin while the rest were resident in other countries. 51 of these users were recruited from the crowdsourcing website

Crowdflower[21], and were required to fill out a survey in English. Those who did not demonstrate an adequate understanding of the survey were disqualified. The following questions were asked of our users. Firstly a series of questions to get an idea of their experience with running and with the Phoenix Park area, around which the experiment was centred and then some questions about the Exercise Builder system.

Table 4.1: Questions asked of users.

1. How often do you run?
2. What is your average running distance?
3. How familiar are you with the Phoenix Park and its popular jogging paths? (1-5) 1: not familiar at all, 5: very familiar
4. Did the website recommend good routes for you? [1 (not at all) .. 5 (very much so)]
5. Did you often alter the recommended routes to your own preferences? Why?
6. How useful were the floating photos ? (1-5)
7. Did seeing the photos cause you to alter the recommended routes? Why?
8. Would you like to use the website in the future? Why?

The participants varied greatly in both their frequency of running and the distance they cover, from some with little running experience to others who run 8km five times a week, with the median being more than once a week for 5.73 km. Figure 4.5 is a breakdown of the relative running abilities of participants, with beginner here indicating less than three kilometre average running distance, intermediate less than eight kilometres no more than twice a week, and advanced meaning greater than 8km or more than once a week (frequent runners and).

Some 20% of participants in our survey were residents of Dublin, but among them 62% indicated they were not familiar with the running routes of the park area. Overall, 69% of those asked indicated they had little to no familiarity with the area in question (see Figure 4.6). The majority of users, 77% (as shown in Figure 4.7), stated they thought the routes that were recommended (prior to altering) were good

[21] http://www.crowdflower.com

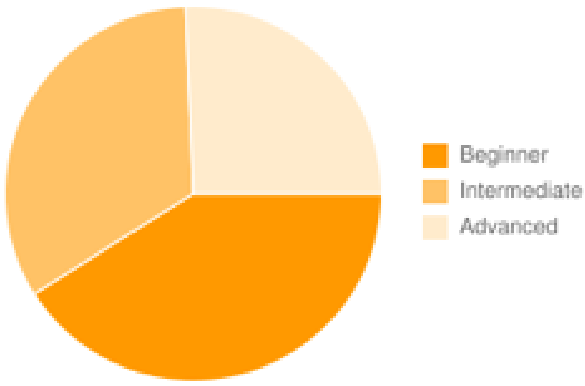

Figure 4.5: Participant experience levels.

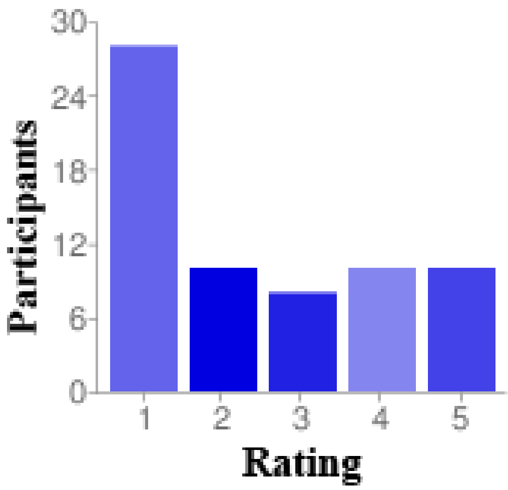

Figure 4.6: Participant familiarity with test area on a scale of 1 (Not familiar) to 5 (Very familiar).

or very good, stating satisfaction with the results of the route combination process. The median score for “Did the website recommend good routes for you?” was 4, indicating a perceived success in traditional recommendation terms. Even still, of the 66 subjects surveyed, only 15 did not alter the recommendation they were given once they specified their distance and starting location, indicating high interest in exploring alternative route options even from a satisfactory baseline.

When asked to rate the usefulness of the photos, the average score given was 4.05 (see Figure 4.8), with users who stated they were more familiar with the park area more likely to give a lower score. The common sentiment across the majority of

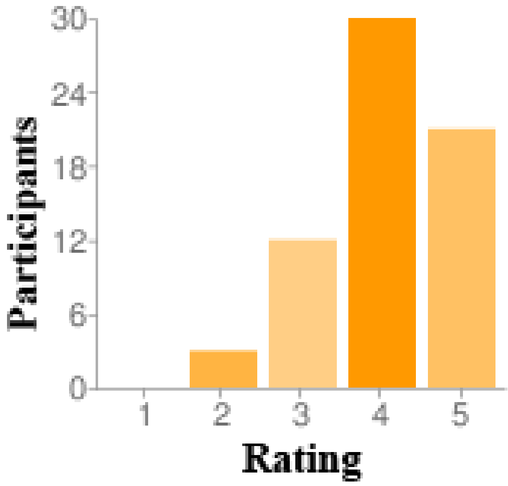

Figure 4.7: Participant rating of initial route quality on a scale of 1 (Poor) to 5 (Very good).

participant comments was that the multimedia data becomes increasingly relevant in areas they have no prior knowledge of, as one subject said they would continue using it "If looking for new runs from an unfamiliar starting point". Other comments highlighted the utility of the photos in recalling areas that have been run in the past, as well as their use as a motivational tool, by picking a landmark or sight outside their usual range for them to run to.

From these results we can see a number of trends. A clear benefit was seen in allowing users to alter their route and, through providing contextual multimedia, inform them of alternative options. This multimedia content was what allowed users with a lack of familiarity about the suitability of alternatives to make informed decisions that ultimately lead to a satisfactory recommendation. Most users (92%) expressed an interest in using the Exercise Builder in the future in order to plan routes in foreign cities or unknown areas, some mentioning they were confident the photos would generate a realistic and useful expectation of the area.

A surprising trend in this data was recognition by the people more familiar with the area of the utility of the embedded multimedia in the interface(shown in Table 4.2). We believe this interest generally extends to any descriptive metadata that

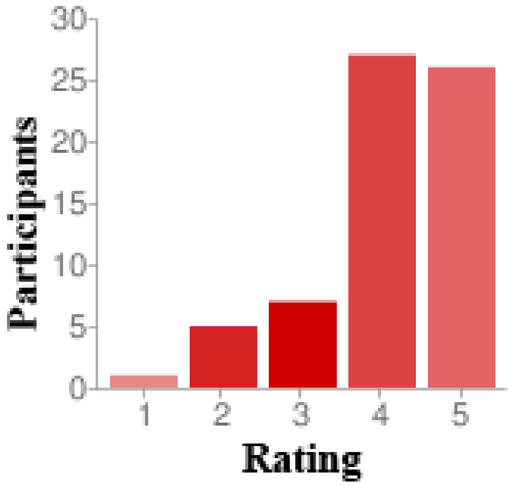

Figure 4.8: Rating distribution of Exercise Builder's multimedia usefulness.

can offer explanation of the surroundings, not simply photos, as comments reflected a general satisfaction with being able to know what to expect.

Table 4.2: Runner familiarity with area and multimedia usefulness.

	Not useful (1-3 on Q6)	Useful (4-5 on Q6)
Unfamiliar (1-3 on Q3)	10	37
Familiar (4-5 on Q3)	3	16

Table 4.3: Exercise expertise and multimedia usefulness.

	Not useful (1-3 on Q6)	Useful (4-5 on Q6)
Beginner	3	24
Intermediate	9	16
Advanced	2	12

It is apparent that usefulness slightly declined with experience (see Table 4.3) possibly owing to experienced runners not requiring new routes. A small group of more experienced runners were less concerned with their surroundings than the additional details such as elevation. This is somewhat in line with Knijnenburg and Willemsen (2009) who showed that users experienced in the recommendation space favour attribute-based preference elicitation, in other words they wanted more control over technical details that lead to their specific recommended run, rather than exploring alternative cases. One other possible reason for this, as one of our subjects mentioned, is that their reasons for not exploring included an existing repertoire of routes they run, “No need as I have my routes and Garmin watch for measurement”. This group were actually less engaged with the photo layer than the users who expressed a higher than average familiarity with the area (many of whom commented that they would have found the photos even more useful in areas they were unfamiliar with).

4.4 Discussion

Here we discuss the findings of our experiment with the Exercise Builder. We have examined user reaction to engaging and revising routes, in order to share better runs and make better runs for themselves. We have been able to study, through a qualitative user survey, the methods that users prefer for interaction with a complex recommender scenario. We found that users thought an interactive approach that allowed them to explore related areas and items rather than simply provide criticism resulted in a more positive recommendation experience.

In this section we presented a method by which multimedia content is used to integrate the user into a conversational recommendation process and the net result of the recommendation is a composite made up of fragments (run segments) of other runs. The Exercise Builder system we built is clearly useful as user interaction can even create new items (run routes) that can be recommended to others in the future. It also showed that multimedia content can assist users in areas where there is not enough information to offer a recommendation. Our plan is to expand the system’s database of routes to cover many other cities in order to allow travellers in new areas. Experienced runners also found features like the elevation graph useful, and contributed to an overall interest in using the application in future.

Our work here builds on work done in interactive recommendation by providing a clear method for users to express their individual interests, context and values. Recommendation commonly has difficulty accounting for a scarcity of initial data on a user, called the cold start problem, but designing for interactivity in this way allows a user more ways to give the system that data, rather than rely on inferred data or ratings, which may take many sessions to build a sufficient model.

Our initial interest in providing an avenue for people new to the area to find suitable running routes proved to be too specific, as more familiar users found great utility in the photos. This information is provided for the purpose of plugging gaps in domain knowledge and providing an explanation of the recommendation. Commercial recommendation is still roughly static, though it recognises the need for recommendation to be viewed as a conversation (e.g. Tunkelang (2011)), this conversation can be part of the design of the system. By taking the view that there are a number of ways users can inform us about the quality of a recommendation we hope to open up recommendation to new areas of study, including interactional context, the idea that context is built and understood through conversation. In general we have found that at any point if we give the user the opportunity to challenge the assumptions the system has made it is to the benefit of both the system and ultimately the user. We have designed a system where a recommendation is a starting point and we give the user the tools to alter it, not only correcting assumptions about them but also creating new route items that can be recommended later.

The novel approach of compound recommendation taken in our example application does not preclude the generalisation of interaction techniques presented here, rather it is shown as a real-world example of recommendations that can be partly altered coherently. Other applications may not be able to create new recommendable items (no new films would be created from an interactive film recommendation session), but this merely means the user's interaction produces a changed recommendation rather than a tweaked or modified one.

Specific areas that could be explored further include how to generalise the approach to other tasks. One clear area of interest is the nearby photo being potentially translated to "related alternatives" in other tasks. This ability to see where an item fits in the overall collection, based on feedback from users, may function as a sort of recommender explanation, which has been shown to improve user satisfaction. In tasks such as movie recommendation the user could be shown "similar to this movie in genre" or "similar and also starring Harrison Ford", to give an idea of similarity along multiple axes. As we begin to explore cases where the resulting recommendation is a compound recommendation from parts of others, as here, there are further design possibilities. These could include blacklisting certain items so they cannot contribute, or weighting how much an item can contribute to the total.

Our findings are indicative of a possible general acceptance of new methods of interactivity in recommendation. It is clear that systems can be designed that allow users to express their will in far more depth per session than ratings or implicit feedback allow. Further exploration is needed to see whether certain recommendable corpuses benefit from specific interaction techniques beyond the general guidelines we have followed. This could easily be generalised to other tasks, to recommend a set of films clustered around an actor or genre, or recommend a set of activities for a holiday or date.

4.5 Comparison to Related Work

4.5.1 Explanation And Knowledge

This work fits squarely within the domain of recommenders that offer explanations. These explanations are said to increase trust in the recommender's offerings, but here we go further by providing context for the recommendation. By being able to visualise the route to run as well as have elevation and photos of the surrounding area the system contextualises *why* the route was chosen. This approach also works for recommendation of sets of items. In our test system the parts (points on a route) only functioned as part of a whole, but the system could equally be used to recommend a playlist of songs with artist metadata surrounding it to let users modify based on artist genre or other information.

Interactive recommendation has been explored in work by McGinty and Smyth (2003), and by Göker and Thompson (2000), but our contribution is an interaction methodology using multimedia content to prompt user exploration. Systems like ExpertClerk by Shimazu (2001) have offered a method to interface with and guide recommendation through acceptance or rejection. Other work such as that by McGinty and Smyth (2003) have experimented successfully with conversation as a method of recommendation also. Such interactive recommendation has generally been used in the past in a number of ways, including as a game to help users discover their true interests in a system, then sharing the information with others, for example by Alon *et al.* (2009). It has also been proposed as a way of preventing information overload in areas such as e-commerce by Shimazu (2001). This approach highlights and then tries to resolve the disparity between a person's actual interest and their perceived interest.

In our work we endeavour to alter the design of recommenders to create direct interaction for a conversational style. Further to this we make compound recommendations based on a large case-base rather than offering single items for acceptance or rejection. Route composition has long been researched, for example by Haigh *et al.* (1997), with commercial applications such as GPS navigation devices being widely available. The frequent area of study in this research is directed toward finding the most efficient, whether in time or fuel or other factors, route from A to B. In our example application we form circular running routes, that finish where they start. This is a route without a target, i.e. it has a start point but runners frequently want to return to that point, which is not a usual task for a route-planning algorithm, so we take a different approach here.

The work presented here is unique because it allows users to capture their unique differences in context and values about what to them makes a good run a priori, something that has not been examined before. As an approach to route-composition it uses crowd sourcing for the purposes of finding routes that are good because of scenic or domain-driven reasons (i.e. frequently run because elevation provides a challenge), rather than optimal in terms of distance or traffic avoidance. The system we have designed and built provides a foothold in the recommendation space, showing users what to expect of the recommended item as well as giving them the means and motivation to find the best run for them. Their expression of uniqueness can be incorporated into the system for future use, expanding the utility of the system. We developed a framework and a demonstrator system, which can solicit useful information from a user and modify the recommendation iteratively after it has initially been made. Our approach overcomes some of the difficulties of information-hungry recommendation approaches, specifically as they relate to new users. The domains that are suitable for this interactive approach are those where the object to be recommended has multiple facets that the user can indicate to be of variable importance, such as choosing a digital camera where cost may be weighted against technical specifications or warranty. Also suitable are domains where the object is a composite of previous recommendations. Further there are design implications and opportunities that may be of importance to future work in the general recommender field. We have chosen the route-planning task for runners or joggers who are unfamiliar with their location, such as holidaymakers or those on business trips to an unfamiliar city.

4.6 Chapter Conclusions and Answer to Research Question

Earlier in this thesis we set out a number of research questions, and the first of those was RQ1, defined below.

RQ 1 How can we create conversational recommenders without intrinsic item knowledge?

In this chapter and the previous one we looked at conversation where no intrinsic item data existed to be discussed, a common scenario. We were motivated to do this by the apparent lack of exploration into collaborative filtering and conversation. We began by exploring the notion of conversation around items. We found that it was possible to offer a conversational experience through discussion of ratings and popularity, the only item knowledge available to collaborative filtering. This conversation had marked improvements over recommendation alone in situations where other conversational approaches could not work. Not only this but we found the relative preferences we collected could be further used to improve recommendations as a source of information similar to relevance feedback, and users found the system a good alternative to Amazon or Netflix-style systems.

We then showed that this notion of conversation without knowledge can lead to interfaces that are by their nature explanatory and interactive. We built a case-based system to recommend compound items, run routes, that provided a unique interactive component. After recommending a reasonable baseline run using the person's needs they are given the opportunity to explore. Metadata that puts the recommendation in context is presented to the user. They are very likely to alter the route to something they find better suited for them, which can then be recorded for later use. Users found this a satisfactory way to approach new areas where they did not have knowledge.

In order to ensure these results were repeatable and verifiable we made use of publicly available data and metrics that described by Herlocker *et al.* (2004). We are the only ones, to our knowledge, to discuss problems related to data collection for recommender purposes, stating it is "very important that the tasks your algorithm is designed to support are similar to the tasks supported by the system from which the data was collected". Because of this we have used a standard recommender dataset made public for the express purpose of testing systems and public run route data made public for users to find new runs.

Having examined these approaches and finding them to be useful we then explored our research question. By making no assumption of knowledge on the part of the user and providing context for the conversation we showed we can create conversational recommenders without intrinsic knowledge. Further work could be carried out to investigate if the variables used to identify an item in conversation, popularity and average rating, could be replaced with other valid variables, including possible metadata. Other vectors of investigation possible would include examining whether a hybrid system that limits the items being traversed by metadata, e.g. only films with the genre "action", would produce an improved recommendation but we will return to this later.

Chapter 5

Information Seeking

5.1 Information Seeking

Having looked at ways to improve the efficiency of conversational recommendation algorithms we now turn to how people use these systems. Computer recommendation is most frequently seen in the form of an informational sidebar in systems, Amazon's "customers who bought this also bought" panel for example, meaning it is passive. As we have shown earlier, we can learn much in terms of information about a person from making the recommendation process interactive and active. In doing so we need to study how users respond to the shift in method, from information absorbing to information seeking.

In this chapter we ask "Do conversational recommenders help fulfil a browsing information need?". Herlocker *et al.* (2004) found "Just Browsing" was a legitimate use of recommenders, elaborating that "Recommenders are usually evaluated based on how well they help the user make a consumption decision [...] we discovered that many of them use the site even when they have no purchase imminent. They find it pleasant to browse. Whether one models this activity as learning or simply as entertainment, it seems that a substantial use of recommenders is simply using them without an ulterior motive. For those cases, the accuracy of algorithms may be less important than the interface, the ease of use, and the level and nature of information provided." However a continuing problem, as mentioned by Ruthven (2008), is that recommenders, e.g. collaborative filtering, push information toward us based on some model of our information preferences, making it hard to define any seeking behaviour. Since people can follow information paths in conversational recommendation by exerting agency (through interaction) can the browsing task be performed with this approach? It has been shown that people wilfully and directly collaborate (Wilson and m. c. schraefel (2009); Evans *et al.* (2010)), so why not blindly, i.e. as in recommendation's wisdom from crowds?

Information seeking (Marchionini (1995)) is the branch of library science concerned with, among other things, how people actually use information retrieval systems to satisfy their information needs. The problem of information overload as discussed by Anand and Mobasher (2005), is not new, and information seeking can be seen as methods used by people to sort, filter, and otherwise make sense of the information that they are exposed to. While this overload occurs in information retrieval generally (Belkin and Croft (1992)), it is the overload seen in recommender systems research (Borchers *et al.* (1998)) that concerns us here.

Together with our algorithmic evaluations in the previous chapter we look here at the perspective of people who use the systems we have built. We do this through the information seeking tools of survey, discussion and, uniquely, detected brain signals. We perform both qualitative and quantitate analysis to investigate peoples' attitudes toward our conversational recommender. We collected their responses to questions and the responses of their brain to recommendations, neural processing which they may not be able to articulate. As noted by Ruthven (2008) "The move from small studies of isolated interactive features to systems that take a more realistic view of how people search is beneficial. A particular theme that has been gaining popularity, and one that has been central to the information seeking literature for some time, is that of task.", leading us to examine users' overall view of the system.

We began by looking at our MovieQuiz application and how people responded to the idea of interacting with a recommendation process rather that querying as in search. It has been shown that convenience is a factor in IS (Connaway *et al.* (2011a)). We here investigate the convenience of conversational recommenders. We surveyed a number of users on their information needs within the system and their reaction to the new approach to recommendation as compared with their everyday life IS (important from a sociology perspective as outlined by Savolainen (2010)) on Amazon or Netflix.

In our brain-scanning experiment we wished to explore sensed signals collected through electroencephalography (EEG). This EEG detects minor electrical currents in the brain which correspond to various signals that tell us about a subject's perspective during an experiment. We looked at sensed responses to recommended stimulus to see if there were any patterns in people's approaches or reactions to recommendation.

5.2 User Study in Conversational Recommender Systems

5.2.1 Approach

As shown in Chapter 3 we examined experimentally the effectiveness of conversation without metadata and some of the potential benefits of such an approach.

After our users (who were described in Section 3.4) had completed their trial use of the system, 33 of the 251 users completed a short questionnaire about their prior usage. Of particular interest in the survey was whether users felt that the interaction improved their ability to find good recommendations and whether users without domain-specific knowledge, or any knowledge of the items they were asked to judge, were at a disadvantage using our system. Previous research (Knijnenburg *et al.* (2011)) has found that users with greater domain knowledge prefer more fine-grained interaction and conversation from their recommender, so we were interested to see if this could be due to other conversational approaches hinging on domain-specific attribute feedback mechanisms such as "Like this but more expensive". In asking the survey questions, of users we were comparing their experience with the system with industry-standard systems from Amazon and Netflix, familiar faces of recommendation in the public eye. The survey included the questions shown in Table 5.1, designed to enquire about users' knowledge levels and their comfort with the system, as a method of finding items, and as a series of questions they could answer easily. Questions one to nine were posed using a 5-point Likert scale. None of the questions had a default answer and all were sent to users remotely and conducted on a webpage.

5.2.2 Evaluation

We recorded some 19,160 activities from the 251 users, with only 381 being "buy" actions (this buy was virtual and no money was involved). The other activities collected were their navigation around the site, their rating, and their relative preference statements. These formed a strong indicator that users used the system to browse the choices provided to them. We found that users who responded to the survey had a wide range of experience and perceived knowledge about movies. The median score for question one, designed to show user experience with the domain area, was 3 on a scale of 1-to-5 (Figure 5.1), with a standard deviation of 1.13, showing that while some were experienced, the average had a casual knowledge on the subject. Question two (Figure 5.2), on the user's own perceived knowledge of film, had a median of 3, with standard deviation of 1.16, indicating that for most movies they had at least some knowledge.

Next we looked at users' acceptance of the recommendations generated, noting that responders found the algorithm recommended fair quality films as shown in Figure 5.4, with one user suggesting a "tag system" be used for genre-specific navigation, i.e. they would like some content-specific features. The majority of users liked the quality of the films, with the number responding 3 out of 5 (I neither find the quality high nor low) roughly in line with the number who didn't feel strongly about their knowledge of movies. Users overall felt that the recommender helped them to discover a reasonably diverse set of films they probably wouldn't have seen otherwise, as shown in Figure 5.5. This is a positive indication that the system helped them find desirable items while browsing.

Finally, we looked at how users found the interface with respect to their recognition of the films offered. Using our approach those asked stated they thought the interface was worthwhile (Figure 5.6). Users had on average an only slightly greater than random chance of recognising films in the system (median score of 3, standard deviation of 0.95), suggesting that in a traditional conversational recommender they would have trouble giving feedback on any item features, and preferred a less interactive approach Knijnenburg *et al.* (2011). However with the approach to conversation we used, users felt that it helped them 'find good items' (Figure 5.7, a median of 4) and even without a high degree of domain knowledge (Figure 5.3, a median of 3) they were able to offer feedback (Figure 5.8, a median of 4). The users preferred our new interface to being offered a list of suggestions (Figure 5.9, a median of 4). This response is a positive indication of their experience with this interface. Finally, we enquired as to how the person found our recommender compared to Amazon or Netflix, the popular retail recommenders. Here the response was encouraging, with many saying the system was well suited for deployment on Amazon or similar, users said things like "Yeah could work on NETFLIX".

Table 5.1: MovieQuiz User Survey Questions

1.	How often do you watch movies, either at home or in the cinema? (Rarely to Daily)
2.	Would you consider yourself knowledgeable about movies? (Not at all to Very much so)
3.	How many of the movies in the system did you recognise? (None to All)
4.	What did you think of the quality of the movies suggested by the system? (Poor to Excellent)
5.	Did you feel the movie recommender offered a good selection of movies you otherwise wouldn't have heard of/seen? (Not at all to Very much so)
6.	What did you think of the "Which do you prefer" interface? (Poor to Excellent)
7.	Do you think the interface helped you find good films? (Not at all to Very much so)
8.	How easy was it to state a preference between two movies in the movie quiz? (Very difficult to Very easy)
9.	Did you find using the interface preferable to just being given a list of suggestions? (Not at all to Very much)
10.	Would you use the interface in future, as part of Netflix or Amazon, as a way to help find movies?
11.	Any other comments?

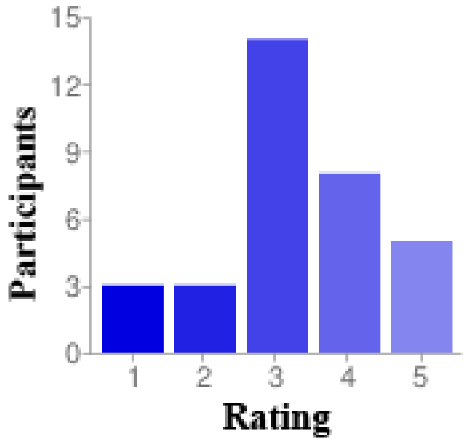

Figure 5.1: How often do you watch movies, either at home or in the cinema?

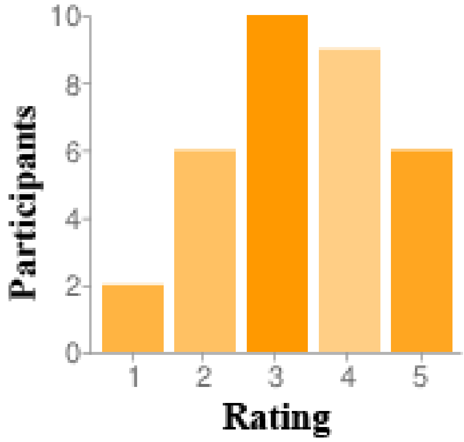

Figure 5.2: Would you consider yourself knowledgeable about movies?

5.2.3 Discussion

Here we looked at how people responded to an interactive interface for recommendations, finding a favourable response. It appears users had no problem stating a preference without domain-specific knowledge, indicating they were easily able to explore the recommendations provided. Since we relied on no metadata or task-dependant information in our work it should generalise to any task a user can judge easily, as in online shopping such as Amazon.

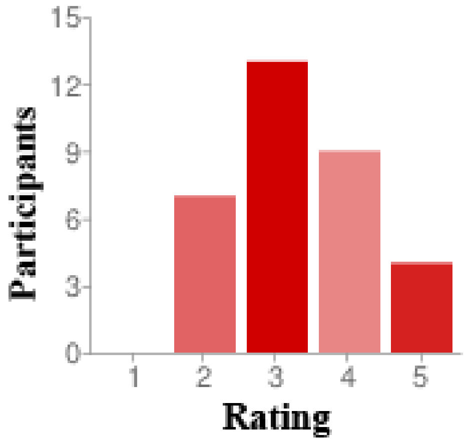

Figure 5.3: How many of the movies in the system did you recognise?

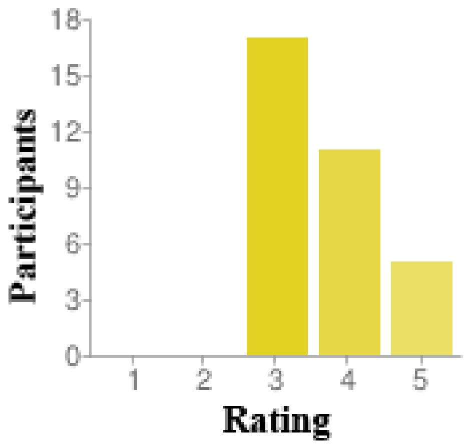

Figure 5.4: What did you think of the quality of the movies suggested by the system ?

5.3 Neural Reactions to Recommended Items

Having looked at how users self-report on their experiences, we wished to look at what we could detect directly in a recommendation scenario. While other information-seeking work has analysed relevance feedback (Belkin (2000)) or even eye-motion (Chen and Pu (2011)) we explored the brain signals linked to information seeking. We decided on the use of electroencephalography (EEG) to detect patterns of brainwaves over a number of other options.

We chose to build on our evaluation of the recommendation approach outlined in Section 4.1, as it suited our task. The signals given off by the brain become increas-

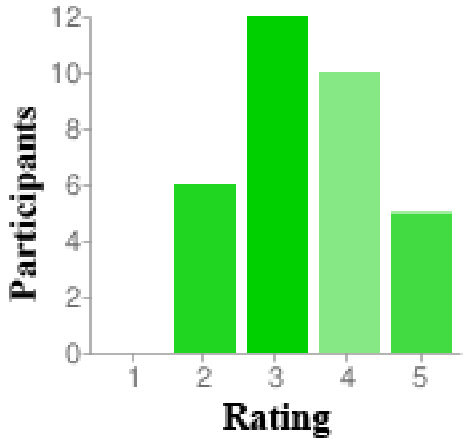

Figure 5.5: Did you feel the movie recommender offered a good selection of movies you otherwise wouldn't have heard of seen ?

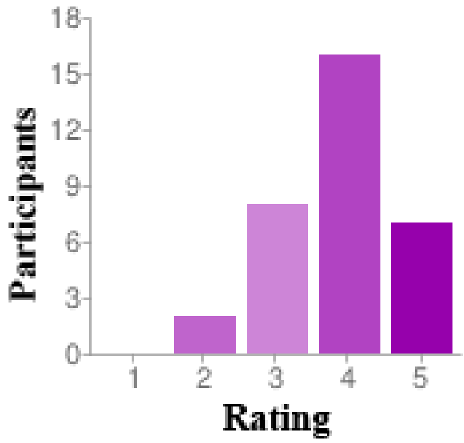

Figure 5.6: What did you think of the interface ?

ingly noisy the longer the subject is required to evaluate an item so the Exercise Builder described in Section 4.2 provided the most suitable recommendable items. Rather than standard text-based information the routes are represented as images, which are the preferred media for EEG analysis, requiring less time to evaluate than text which results in a cleaner signal. Secondly, and also a consideration, is that the test subjects had a familiarity with the Dublin area, allowing them to interpret and respond to images quickly.

Recent research has shown that presenting the results of a recommendation process as a conversation between user and system has merit. User feedback, collected

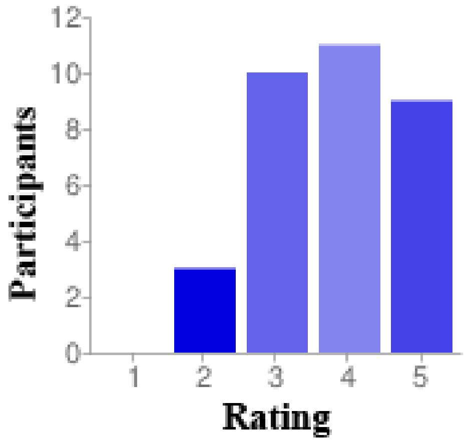

Figure 5.7: Do you think the interface help you find good films ?

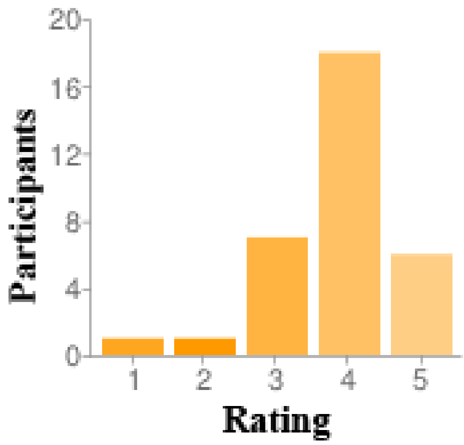

Figure 5.8: How easy was it to state a preference between two movies in the movie quiz ?

in order to build a user profile on which to base recommendations, can either be *implicit* and based on user behaviour, or it can be *explicit* where a user directly feeds back judgement on the relevance of objects to be recommended. Explicit user feedback can be time-consuming and puts cognitive load and stress on the user, even when framed as a conversation.

Recommender systems are fundamentally limited by the degree to which they can measure the relevance of the items that they recommend to a user. These systems involve training and customisation, relying on feedback provided by the user which determine the performance and quality of the recommendations. Requesting a user

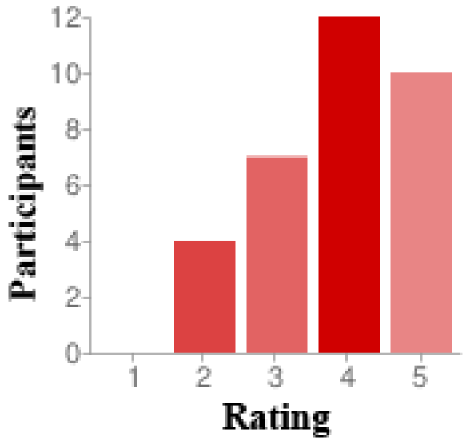

Figure 5.9: Did you find using the interface preferable to just being given a list of suggestions ?

rank or rate an item's relevance is the usual method employed to gain perspective on the user's preferences. As the number of these ratings increase a recommender algorithm more effectively converges on future recommendations that will satisfy the user's recommendation needs.

Although users may *explicitly* indicate preference / non-preference of a selected item, *implicit* indicators are often more telling. Work in music recommendation shows that users may rate items highly while actually preferring others, behaviour that is revealed through implicit feedback such as song play counts. To this end we explore a user's brain responses to images recommended to them. By observing EEG signals, explicit feedback signals from the brain, present during the acceptance or rejection of recommended items we may be better able to unveil the mechanisms that determine choices and preferences, so as to ultimately improve these systems, allowing them to better serve the user. Our work here endeavours to explore how the brain reacts and to see if there are any clear implications for recommendation.

In this work we explore the role of visual multimedia in both explicit and implicit feedback. We validated our approach to explicit feedback with a route recommender. Following this we looked at implicit feedback in an application for route planning for the routes recommended which are represented as images, directly monitoring the response from EEG signals while users rated items. What we find is that in certain instances patterns of activity are present prior to the presentation of the recommendation, general to the task of being recommended, that indicate a bias before the user ever sees their recommendation.

5.3.1 Approach

EEG measures voltage fluctuations resulting from ionic current flows within the neurons of the brain. It is preferred to other methods of neural signal measurement such as functional magnetic resonance imaging or direct sensing of signals on the surface of the brain - accomplished through surgical procedure - because it is easier to deploy, less invasive and less costly. A number of electrodes placed on the scalp are used to measure the voltage fluctuations resulting from ongoing brain operation. The number and placement of these varies based on task (work such as that by Healy and Smeaton (2011) is being done to make it easier to set up and even less invasive), but 6, 16 or 32 are typical.

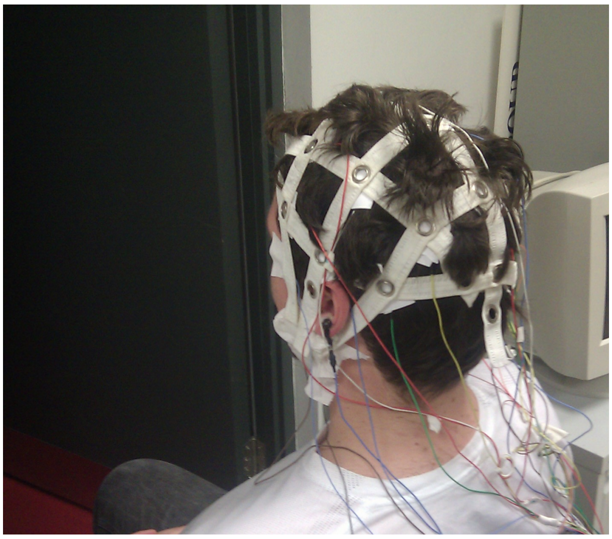

Figure 5.10: A typical EEG setup using 16 electrodes. In our work we use six.

This multiple electrode setup allows the detection of EEG signals that have complex spatial and temporal profiles in how they appear on the scalp at across time. By examining how these signals change in relation to an event such as a stimulus, we can detect something about a user's brain reaction to that event. This is known as the Event-Related Potential (ERP) technique. Due to the low signal to noise ratio of these signals a number of signal time windows (epochs) of the same event occurring are often recorded, and then averaged in order to mitigate the noise and expose an underlying common signal component of interest - these are called ERP components. This process often reveals differences in amplitude or timing of these components in relation to various stimulus conditions.

A number of these ERP components are typically seen when dealing with visual stimuli. These components are labelled based on the direction of voltage deflection (positive/negative) and a shortened timestamp of when it typically occurs, such as P1, N1 (both happen around 100 milliseconds) - both of which are typically elicited in response to a visual stimulus. Other ERP components like P300 (300 milliseconds) are known to co-occur with visual stimuli that capture a subject's attention in a significant way. These signals and components differ in latency and amplitude from task to task and from person to person, and have been the subject of much study.

In this thesis we focused on using FFT (Fast Fourier Transform) analysis in order to reveal frequency oscillations around events. Numerous approaches in analysing these signals with both time-domain and frequency-domain representation have shown that each approach offers information the other does not. These signal transforms have previously been used in situations say to reveal and detect dynamics of signals during real or imagined movement so as to drive a brain-computer interface and more recently (Healy (2012)) as a way to analyse timed event related signals.

This makes clear signals that differ between different stimuli, helping to highlight the cognitive processes that lead to them. We used FFT because it mitigates some issues present with traditional methods, in that it can reveal subtle frequency changes in brain signals. From this we are able to compare the collected epochs of different events as a means of seeing what is unique about an event.

We then took all the instances of all stimuli for each single person and trained a Support Vector Machine (SVM) for each set of two different stimuli conditions we wished to compare. SVMs (Chang and Lin (2011)) are supervised learning models used in data analytics for the purposes of classification and pattern recognition. Using them allows us to detect if signals are present whenever a person sees a stimulus comparatively to another stimulus.

We generated a dataset of 198 images recommended for our Exercise Builder, described in detail in Section 4. These images (such as in Figure 5.11) were flipped in order to prevent recognition of a face shape present in a road through the Phoenix Park, which would have generated noise in the EEG signals.

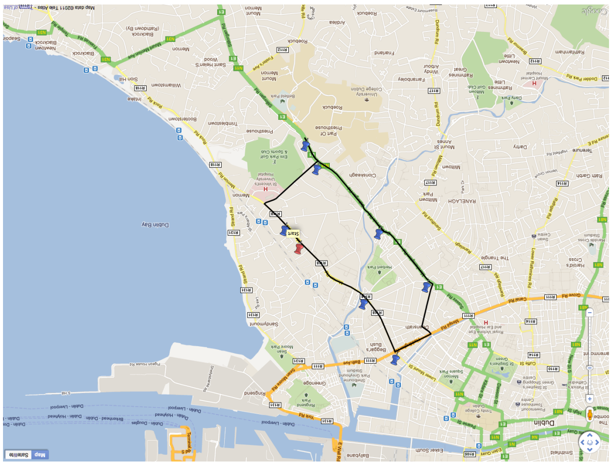

Figure 5.11: An example recommendation from our brain dataset

5.3.2 Evaluation - EEG Analysis

We considered that since users found it easy to decide on good routes perhaps it would be possible to detect this implicitly. We used the EEG signals to study brain activity relating to seeing recommendations, that were then accepted or rejected. In order to simplify the task (as complex tasks such as reading, rating or interaction would lead to more noise in the signal) we used recommended images. Images are quickly evaluated and are common in EEG tasks such as target search, making them well-understood test data. We generated a corpus of images that featured running routes from our Exercise Builder (detailed in Section 4.1).

The dataset consisted of 198 images, randomly generated from the system, 66 images for each of three categories; recommended images, interacted with recommendations and control images. The 66 recommended images were built through the recommendation algorithm, providing it with random starting points and distances in the range of 2-10km. The 66 images grouped as "interacted with recommendations" were made in the same way, but were each interacted with by stating a preference for distance over popularity, altering the recommendation. The last group, the control, is 66 of the routes that form the case-base which were actually run by someone.

To determine whether patterns of differentiating EEG activity existed prior to the presentation of a recommendation we used a machine learning analysis on the EEG signals. EEG signals were recorded using a KT88-1016 amplifier system. Electrodes were placed at 6 locations Fz, Cz, Pz, Oz, P3, P4 as per the international 10-20 system placement map (see 5.12 [22]). The left earlobe was used as a reference with the chin as ground. Signals were digitised at 100Hz and subsequently band-passed from 0.1Hz to 15Hz.

Epochs of the EEG signal from the 6 channels were extracted corresponding to the 7 seconds before the recommendation image was presented on screen, and were subsequently labelled as Control, Interacted With, Recommended. Using an FFT algorithm [23] we extracted features from the frequency domain (0Hz-15Hz) of each of these epochs using multiple time periods, beginning 7 seconds prior to the presentation of the image and incrementing in 1 second intervals towards it. A feature vector was formed by amalgamating the features derived for one time period across

[22] Source: `http://www.immrama.org/eeg/electrode.html`
[23] NumPy - `http://numpy.scipy.org`

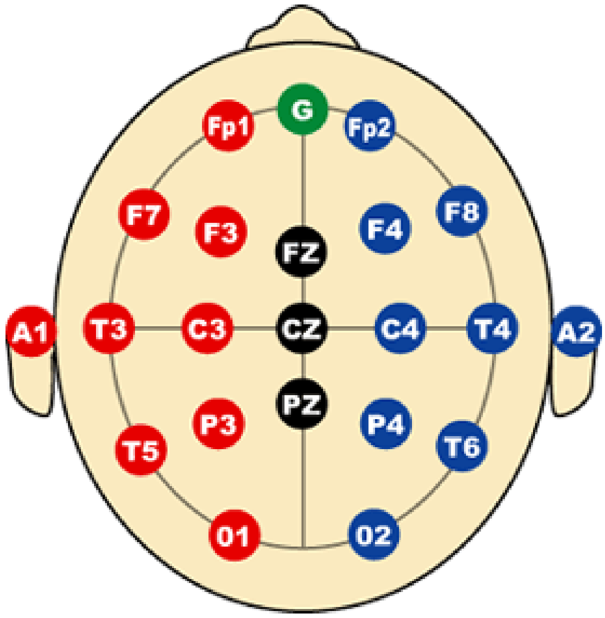

Figure 5.12: EEG node placement diagram.

all channels. Doing this across all 6 channels yielded a feature vector 90 elements in length. Each of these feature vectors was further labelled as corresponding to an accept or reject. These would be used to determine if there are significant similarities or dissimilarities across users' brain signals.

We used a repeated random sub-sampling cross validation approach with a PLS Regression algorithm (Wold *et al.* (2001)) in order to assess the level to which signals occurring prior to the each image presentation displayed indicators determining the selected response. Testing sets were comprised of 5 randomly-sampled examples from each of the cases to be compared, with the remaining examples used for training by the PLS Regression algorithm. Reduced numbers of training samples might be the cause of low classification accuracy in certain instances. The validation procedure was repeated 100 times with randomly sampled training and testing sets. On each iteration of the validation procedure we used the trained model to generate predictions for the examples in the testing set. By demonstrating that the trained model is capable of doing this above the likelihood of chance we can assert the presence of discriminative information present in the signals that allows them to be differentiated. We used a measure derived from ROC-AUC (Fawcett (2004)) where we calculated accuracy as $|AUC - 0.5| * 2$. Accuracies across all iterations were then averaged to give an overall score.

To assess the significance of measures yielded by this validation approach we used a bootstrap resampling method and we repeated the entire validation process 1000 times randomising the labels in the test set on each iteration. We calculated the probability of an accuracy exceeding .064 as being less than 1/1000 (p=.001).

5.3.3 Experimental Results

While a number of behaviours can be seen in the data we recorded some particular elements are of note. We found that in some instances the timeframe immediately prior to some users' acceptance or rejection of recommended media showed indicators to say they had already made their decision. This is a significant finding as it shows that regardless of the subject-matter the user was presented with, their resulting rating in said instances is not representative of an unbiased evaluation of the material.

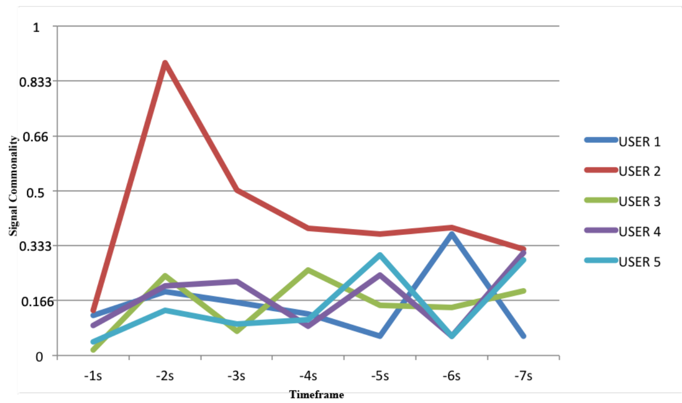

Figure 5.13: EEG difference between accept/reject signals in Recommended items.

In Figure 5.13, 5.14, 5.15 we show the accuracies derived using the outlined machine learning method on signals from the 3 recommendation cases (Recommended, Control and Interacted With respectively). For each of these we can see the differ-

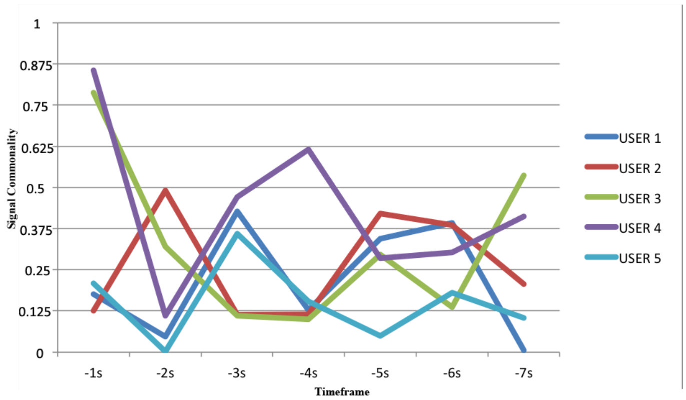

Figure 5.14: EEG difference between accept/reject signals in Control items.

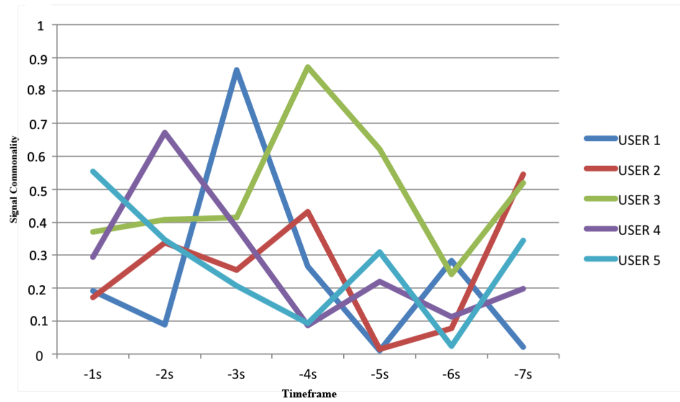

Figure 5.15: EEG difference between accept/reject signals in Interacted With items.

entiability of the EEG signals occurring prior to the image presentation that predict an accept vs reject for that image. It's evident that for many users in a number of instances there is a detectable indication prior to acception/rejection. There are patterns of highlighted differentiating activity present in the frames directly prior to the subject providing their response.

The presence of activity like this might well indicate that recommender systems can be fundamentally limited by the degree to which they can measure the relevance of the items they recommend to a user. That is to say that user rating can be unreliable in the presence of this activity. Conversely, we might argue armed with this knowledge we could better utilise susceptive states in order to maximise the probability of acceptance, or a truly considered rating, of a particular class of items.

Users approach expressing preference in different ways. For example in Tables 5.2, 5.3 and 5.4 we can see that although some overall similarities exist between users, there are important differences present, such as a bias towards selecting accept over reject overall. This is interesting, especially since it can be seen that recommended items are generally more accepted than control, a notable difference we cannot explain. Utilising EEG surrounding such events can allow us insights into the mechanisms involved in generating these behaviours so that we may better understand the reasons for them and their implications.Where a frame had 2 or more responses we discounted the answer, e.g. User 1 had frequent multiple responses.

5.3.4 Discussion

In this section we explore how EEG signals present during the recommendation process can assist us in understanding a user's choices. Our work brings us closer to discerning the reasons behind human choice, especially in terms of pre-recommendation predisposition.

We demonstrate here how EEG signals may be indicative of the quality of a user's evaluation of an item and may highlight predispositions. We showed that patterns of similarity exists between the users in the frequency of their selections across the conditions. Even in a basic recommendation task, a diversity of user preferences and strategies are present. In this work we have examined the use of multimedia in explicit and implicit feedback. We found with explicit feedback that users are easily able to alter suggestions when given the means to form reasons to do so. We also found that there exist signals detectable in the brain that indicate a bias in subjects before they have even received their recommendation. These signals could be interpreted twofold; as the user's recommendation in particular instances not being representative of their opinion, as there is evidence of a pre-biased state

Table 5.2: User responses to control route image stimulus

User	Accept	Reject
User 1	30	28
User 2	45	22
User 3	36	27
User 4	32	35
User 5	27	40

Table 5.3: User responses to recommended route image stimulus

User	Accept	Reject
User 1	51	7
User 2	54	12
User 3	51	11
User 4	38	27
User 5	46	20

Table 5.4: User responses to interacted recommended route image stimulus

User	Accept	Reject
User 1	51	12
User 2	41	25
User 3	35	29
User 4	23	43
User 5	26	40

regarding the response selection process, or that we could perhaps bring greater user satisfaction by strategically recommending items at the most opportune moments. Future work can be done to further explore the question of the best approach to the brain activity discovered here.

5.4 Comparison to Related Work

5.4.1 Information Seeking in Conversational Recommendation

We have previously discussed related work in information seeking in Section 2.1, here we contextualise our contribution by specifically highlighting how it adds to the existing body of work. Recent work (McNee *et al.* (2006)) has explored ways to "act as a bridge between user information seeking tasks and recommender algorithms.", describing the dialogue, the recommender's personality and the user's information seeking need as key pillars in this task. Our work in this area contributes to an exploration of recommender systems whose dialogue, a key part in the "Just Browsing" task, is an explicit conversation.

It has been shown (Morris *et al.* (2010)) that users can get help from their friends to fulfil information needs, and even that recommenders can help build queries (Belkin (2000)), but here we present exploratory work on the role conversational recommenders can play in information seeking. Social influences have been seen to play a part in IS strategies Allen (2003) and recommenders have long been used to aid these social strategies McDonald and Ackerman (2000a) but here we explore direct interactive recommendation as a solution to information needs.

The existence of exploratory search has also been reflected in research with methods to explore and build queries using browsing as feedback, both for conventional information retrieval (Salton and Buckley (1990)) and multimedia (Campbell (2000)).

Our work here takes this approach to browsing to build a current context or profile for recommendation along with the user's conventional recommender profile.

5.4.2 EEG Experimentation and the Examination of Recommendation Response

There are two common methods by which recommendations are formed Adomavicius and Tuzhilin (2005). *Content-based* and *collaborative* systems are both generally designed to become more powerful and accurate as usage increases. This is a limiting factor as it can adversely affect recommendations for new users or recommendations for new items, as a lack of information results in poor grouping. Our work here uses multimedia in a descriptive role, not prompting feedback. Hybrid systems have been the common approach to solve this. We show here a system for recommending routes from a case-base and collecting feedback from multimedia related to the area around the route. Route recommendation is well-researched (McGinty and Smyth (2001)), but here it is used merely as an example to examine the effect multimedia can have on explicit feedback, post-recommendation. To assess the impact of this we explored the naturally-occurring spontaneous reaction of the brain to images generated through this recommendation, in order to see if a noticeable pattern of acceptance or rejection could be detected.

Implicit feedback has been long been used to augment a system's understanding of a user, by collecting behaviour data from sources such as page-views or directly from sensors. In this work we examined the EEG (electroencephalography) signals coming directly from the brain as users assess recommended media representing routes generated through recommendation and including interaction. EEG signals can be detected on the scalp by measuring voltage fluctuations between points using conductive electrodes. These signals are known to be reflective of cognitive processes such as those involved with attention and decision making. There has been a recent trend of using signals of this type to allow communication between user and computer Shenoy and Tan (2008). One example of this for instance is an application wherein a user engaged in a search task may be assisted in converging on relevant search items by analysing responses in their EEG signals following the presentation of each item Pohlmeyer *et al.* (2011). In this section we use analysis of EEG signals to better understand the processes that contribute to the acceptance or rejection of a recommended item. In particular we are interested in patterns of EEG activity prior to the presentation of a recommendation, and how these might indicate whether a recommendation will be ultimately accepted or rejected.

5.5 Chapter Conclusions and Answer to Research Question

RQ 2 Do conversational recommenders help fulfil a browsing information need?

In this chapter we looked at how people engage with conversational recommenders. We were motivated to do this by the shift from passive recommendation to seeking afforded by the conversational approach, that has not been studied. We found that users enjoy being able to browse recommendations even when their knowledge of the item domain is low.

We expanded on our examination of user satisfaction with interactive recommenders as shown in Section 4.3 by surveying users of our MovieQuiz system on their interest in conversational recommendation. This study of users was observational in nature but points to an interest in examining items they have no experience of, in essence using the conversational approach as a learning experience. This builds on our conclusions in Section 4.6, where we found users with little domain knowledge could use conversational recommenders unhindered if the approach acknowledged this, in that these less experienced people used the conversation to gain confidence and experience to make decisions. This was also evident in Section 4.3 where users mentioned the conversation helped them get a feel for "the lay of the land", an understanding of the context needed to make a decision.

We found that users of conversational recommenders approach the system as a method to browse item choices. This indicates that conversational recommenders are well-suited to the "Just Browsing" task as outlined by Herlocker *et al.* (2004). Our study of the brain's reaction to recommendations indicates that there are detectable signals which show a user will reject what they see next. These signals could be present for a number of reasons but represent a new source of information and insight with regard to recommendation and serendipitous discovery, in that perhaps a recommendation failed or was dismissed because it was recommended *at the wrong time.*

Chapter 6

Social Context

6.1 Social Context

In the previous chapters we looked at ways in which we could expand the compatibility and overlap between conversation and recommendation, and we also examined how users would react to systems with such hybrid interaction. We next turn our attention to other factors that could contribute to the interplay between conversation and recommendation. Since a fundamental aspect of recommendation is grouping people by the things they like, we have focused on other groups that people *choose* to become a part of, that is their social relationships. Social connections have become a key point in the discussion of recommendation (as shown by Mobasher *et al.* (2012); He and Chu (2010a); McDonald (2003)) and in this chapter we will present a set of experiments where we examine social relationships and how they may be useful for recommendation purposes.

Since the development of recommendation as a form of information retrieval, it has usually incorporated a social aspect (clustering people or items based on people's feelings Schafer *et al.* (2007) for example), building on the information provided by others to generate "word-of-mouth" suggestions. Recommendation methodologies have grown to the point that now we are able to use information about things to more accurately suggest items when other people aren't available for social input, but social uses of computers have also grown. Work done by Abdul-Rahman and Hailes (2000) showed that in contexts such as around movies or other recommendable items, people develop social connections with others who have similar tastes, work that was later developed in Ziegler and Lausen (2004). Social uses of the Internet have developed, with people now writing blogs to share reviews, posting to their network of friends on sites like Facebook, or putting short messages on Twitter. Since these new channels encourage people to share their opinions on a wide variety of topics, including products and services, much attention in social recommendation has been given to scraping these sources for sentiment or activity around recommendable items, as well as for measuring the trustworthiness of sources (O'Donovan and Smyth (2005)).

Our interest, as it relates to our research question below is the relationship between people, their friends, and the larger social circle they may have an affect on. We are interested in how social relationships affect the judgement a user makes on an item. The hypothesis behind our research question "Can social relationships inferred from contextual cues prove useful in improving recommendation accuracy?" is that our friends and associates exert a degree of social influence that can be detected and used to compute suggestions for items. That is to say, we do not judge things in an objective vacuum based solely on their merit, they exist in a context within our lives, so comedies always seem more enjoyable with a group of friends than viewed alone. It has been shown that emotional connection is hugely important to how we judge items or experiences, not only subjectively but also objectively. The Significant Objects[24] project examined the monetary effect of adding a (fictional) emotional story to items purchased at a thrift store and sold on eBay. They found[25] that objects sold for far greater than cost price with their emotive stories tied to

[24] `http://significantobjects.com/about/`
[25] `http://significantobjects.com/experimental-results/`

them, showing the objective influence of emotion. Seeing others share reviews of things you have yet to experience may form its own emotive bias in a person's mind, that we are interested in exploring. Other work (Koren and Sill (2011)) has looked at how people rate items and shown that ratings are subjective and meant to order from "most favourite" to "least favourite" rather than 4 stars being exactly 1 unit of measurement better than 3 stars. We are interested in seeing if that subjective experience, which leads to objective changes in rating, can be influenced by detectable factors such as what your friends have previously publicly said about an item.

6.2 Social Dataset

We collected a dataset for our purposes from the book-collection website Goodreads[26], from random users and items that appeared on the site's frontpage. Goodreads is a site that allows users to share ratings and reviews of books with friends, and functions as a social hub around books in many ways. For our purposes it also makes an interesting semantic distinction between users who are regular readers and users who have written books in their collection, i.e. authors of published books. This acknowledgment that creators of books also contribute reviews is interesting as it allows us to examine whether these creators could or should be considered expert, trustworthy or useful for recommendation purposes. Our experiments detailed in this chapter investigate how social relationships can influence recommendation. We show how this data could be beneficial when integrated into a conversational recommender. While much work has been done on social recommendation our unique contributions are to analyse the effects of differing types of social relationship and how they may be beneficial.

We gathered the data detailed in Tables 6.1, 6.2 and 6.3 from Goodreads. In

[26] `http://goodreads.com`

order to do this we downloaded all of the users linked to on the front- page of the site. We then downloaded all their reviews, and all their friend connections, then used these connections to download more users, whose reviews we also downloaded. Some of these users were annotated as author and we recorded this. This continued recursively until we were left with the dataset described. We then downloaded all the information for all the books reviewed.

Table 6.1: Rating statistics for the Goodreads dataset.

ratings including "to read" items	161,237
actual ratings (not "to read")	95,307
average number of ratings per user	46.58
average number of actual ratings	23.32
users with ratings	3890
users with actual ratings	3648
authors who are Goodreads users with reviews	2747
non-author users with ratings	1181

6.3 Social Trail

In order to explore the concept of social context we looked at how users are influenced by other people. We studied the effects of a person sharing their opinion of a book on Goodreads on the expressed opinions of people who saw it. Our interest was in detecting an actual social relationship of influence in these "trails" from one user to another. It has already been shown (by Groh and Ehmig (2007)) that directly connected friends tend to have similar opinions, but we extended our examination

Table 6.2: Miscellaneous statistics for the Goodreads dataset

total user profiles, with currently reading books	4,382
friendship relations	846,682
books	28,599
books by "user" authors	7,163
total reviews	158,899
total actual reviews	35,348
reviews that say "recommend"	3,591

Table 6.3: Example Goodreads rating details

Collected Information	Example Value
User ID	2147919
Book ID	7604
Rating	0
Review ID	49941962
Average Rating for this item	3.78
Author ID	5152
Rating Added	Sat Mar 21 05:11:23 -0700 2009
Rating Updated	Sat Mar 21 05:11:23 -0700 2009

to look at how apparently unconnected people could be seen to influence each other through third-party friends. It has also been shown (by He and Chu (2011)) that there are a number of social issues that confound traditional recommenders, such as being mislead by friends, which is one reason why we hoped to examine complex social relationships. This contextually sensed complex social relationship was then examined to see if it could be exploited to improve recommendation accuracy.

Trust as it is called in recommendation is an attempt at "defining the goodness of a user's contribution to the computation of recommendations"(O'Donovan and Smyth (2005)). In this work the term trust does not really suit, as we are not defining an objective "utility-in-recommendation" value, though in concept our approach is similar. We wish to define how much all other users who are in any way connected to a given person, will influence that person's ratings in order to account for that influence in the recommendation. This is a distinct social context because it is unique to that user, drawing on the collection of others connected to them, whose prior expressed opinions agreed with them. In this way it could be said to be a measure of one user's trust in others who are socially connected to themselves, a peer-to-peer reputation rating or perhaps more clearly how much others can be said to predict the user's rating. To avoid ambiguity we will not refer to our approach as trust-based recommendation, though it is influenced by it.

Our examination looks at common deviation from the mean score given to an item among users with some connection. This measure is ordered by time, allowing us to see which people have ratings that predict a user's own ratings. This can be seen as a measure of potential influence, although the temporally-ordered correlation of agreement beyond the mean does not prove (or disprove) one user directly influencing the other, but it does indicate a subset of users who share similar opinions to the user, before the user. This subset includes people who legitimately directly influence, distantly connected influencers such as trend-setters and people who could be said to influence only by having expressed similar opinions earlier than the user (who they have some social connection to). This conflates a number of signals under the banner of influence in order to examine whether they have a detectable impact and use.

Another possible measure of this sort of potential influence would be traditional recommendation metrics such as mean absolute error or RMSE, measuring the difference between expected rating and actual rating with respect to other users. However these measures would convolute the impact one user has on another with the error figures of whatever recommender system was used. For this reason we solely analyse the data gathered without a recommender system.

6.3.1 Examining Social Influences

We examined the rating habits of users across the collected Goodreads dataset for indicators of important social relationships. Following our hypothesis stated earlier, we were interested in relationships that resulted in an influence to the user's rating of an item. Social recommendation frequently looks at scraping data or sentiment from social sources such as Twitter[27]. Here, though, we wish to study the actual effect of one user expressing their opinion on other users in the system with whom they may or may not be friends with.

In order to look at the influence of one user on another we looked at the difference

[27] http://twitter.com

between their shared opinion and the average opinion on the book computed over everyone who has read that book on Goodreads. This comparison of local average to global average is intended to show how much one person's opinion might predict another's, or how much people agree with each other in differing from popular opinion. Our hypothesis here was that if one user expressed their opinion before the other, that user could be said to predict or influence the other, or at least inform us as to what sort of opinion the other will have.

We looked at the 161,237 ratings on a five-star scale in the Goodreads dataset in an attempt to track influence. This includes some 65,930 "ratings" which were markers of intention to read in future, marked by the system as 0 ratings. Theses were included since they still may represent an influential or predictive relationship, in that one person plans to read a book because they heard about it from a friend, or saw their friend was planning on reading it.

Our first interest was close social influence, the influence of a person on that user's list of friends. We recorded the relationships between people based on common books they reviewed, with the relation recorded as time-based, i.e. the person who first posted a rating was deemed the influencer. We then looked at less direct or obvious influence present in the data, examining people not in a person's social circle but who may feel influence from their opinions. Our theory was that while a user expressing an opinion (an actor) might not be seen directly by many others, their action might have some trickle-down effect on others (affected parties). For each person in the collection we looked at their effect on people who weren't on their friends list but were connected to some degree through those friends on the user's social graph.

We further were interested in whether there was a difference between positive and negative sentiment in terms of effect. In order to show sentiment we separated the cases of influence into positive and negative differences from the global average, counting cases where both users have marked an item as "to-read" as positive incidents of pre-experience influence, i.e. the influencer's opinion on what could be good is being trusted by the subject of influence.

6.3.2 Weighting Social Influences

Having analysed the observed effect of users sharing their opinion we wished to investigate the best method of integrating this knowledge into a recommender system. In order to improve the accuracy of item suggestion to a user we wanted to account for not only the type of user they are (as in regular user-based collaborative filtering) but also whose opinions affect them. We set up a user-based collaborative filtering algorithm using Pearson correlation to determine user similarity as a baseline, and for each approach performed five-fold cross-validation to arrive at our final figures.

Since close social ties proved to be the most easily detected they were the first to be used to improve recommender accuracy. We developed a function to increase the weight of influential peoples' opinions on those they are seen to influence within the system as shown in Algorithm 2. This function allows us to add the social cues we have detected to the traditional collaborative filtering algorithms. The weighting was cumulative; a user's weight was the sum of all the instances they shared of notable opinions differing from the mean. This weighting was added to the originally computed similarity, it did not replace it. This meant a stronger weight was given to someone who continually had a similar opinion to a user prior to that user. The average weight (i.e. sum total deviation from the mean by one user on another) over both close and distant instances in the the data was 1.79. This would indicate that, as is generally assumed in recommendation, friends do not necessarily have more in common than strangers (as this commonality is based on the number of shared books rated).

Equally we wanted to examine the part distant influences present in the dataset might play in improving item suggestion for users. Just as user-based collaborative filtering will group disparate people based on their interests, we wish to explore whether distant social influence represents another source of potential connection to improve recommendations. In Algorithm 3 we present the algorithm with which we constructed a weighting matrix for distant social ties.

For both close and distant apparent influences we filtered the cases that were less than one star from the global average to give a list of the 'notable' social influences on users. We intended to examine these large differences from the mean to see if they could be reliably used to indicate items of interest to a person. These influences were then integrated into a recommender system.

We implemented a standard user-based collaborative filtering algorithm as a control, against which to test our socially-weighted variations, using Pearson correlation to determine user similarity. For each test we withheld a percentage of the ratings (20,40,60 and 80 respectively) in the collection (all the users, including authors, but only the 95k actual ratings were used) to see how well each approach could predict them, as indicated in the tables. We generated an unbounded list of recommendations for each user. Our social weighting algorithms in Algorithm 2 and Algorithm 3 were used to generate weights that were applied during the recommendation to indicate users who have a perceived influence or correlating opinion, are seen as more similar than they usually would. In our experiments "weightval" was set to 0.2 to push similarity closer at each similar social connection.

6.3.3 Results

We began with our analysis of the dataset. Among the 3,890 users present there was evidence of 1,528 unique users originating 58,685 different instances of common deviation from the mean. In Table 6.4 we categorise these instances, showing that more than half were between directly connected friends. We classed each instance as either positive or negative based on whether it resulted in a positive or negative

Algorithm 2 Close Social Score Weight Generation Algorithm

$weightval \leftarrow weighttobeaddedpercommonality$
for all $user$ **do**
  for all $friend$ of $user$ **do**
    if $(friendRating + userRating)/2 - averageRating \geq 1$ **then**
      $Weights[friend][user] += weightval$
    else if $friendRating + userRating)/2 - averageRating \leq -1$ **then**
      $Weights[friend][user] -= weightval$
    end if
  end for
end for
return $Weights$

Algorithm 3 Distant Social Score Weight Generation Algorithm

$weightval \leftarrow weighttobeaddedpercommonality$
for all $user$ **do**
  for all $stranger$ to $user$ **do** ▷ If the stranger is a friend of a friend (to any degree)
    if $stranger \in usersSocialGraph$ **then**
      if $(strangerRating + userRating)/2 - averageRating \geq 1$ **then**
        $Weights[stranger][user] += weightval$
      else if $(strangerRating + userRating)/2 - averageRating \leq -1$ **then**
        $Weights[stranger][user] -= weightval$
      end if
    end if
  end for
end for
return $Weights$

difference from the mean, i.e. how the two people felt that was generally different. In order to see what proportion of these were significant we filtered these items to show the "notable" influences that resulted in a difference of at least one star between the local average and the global average. Lastly we found that this could be split between 7,588 actual ratings and 5,279 cases of both users giving "to-read" markers. This shows a roughly equal split of effect between "I agree, I should read this book too" and "we both feel similarly about this book", or pre-experience and post-experience effect. This is interesting in that pre-experience effect is distinct from any knowledge of the item aside from a blurb and cover that appears on the site, while post-experience the affected party has had a chance to develop their own experience with the item, enough to rate it.

This yielded, as seen in Table 6.4 36,641 close incidents of influence to any degree between friends. It further showed that 12,867 of these cases of close influence were notably high, with 7,588 being indicators after both actor and affected party have had experience enough to rate. We found 22,044 cases of an actor expressing a common opinion prior to a possibly affected party that they had loose social ties to. Of these 6,960 could be said to be notable, constituting at least a 1 star difference from the mean, and the split between pre and post experience influence was 3,404 to 3,556. We see that while the numbers are diminished in comparison to the close ties there are still roughly equal ratios of effect numbers. It is especially interesting to note the "to-read" zero-score markers make up roughly half of the distant notable numbers, in line with the ratio of "to-read" to rated in the close numbers despite a lack of contact between the users, suggesting either trickle-down influence or notable commonality between the users.

We saw a much greater number of both total and notable positive connections, shown in Table 6.4. However, once both parties had experience of the item and actually rated it (as shown in the "Notable Non-Zero" column), a negative connection was more likely. This appears to show that people will readily take suggestions on

Table 6.4: Rating Sentiment Influence on Goodreads dataset

Category	**Total**	**Notable**	**Notable Non-Zero**
Close Positive	25,347	7,316	2,037
Close Negative	11,294	5,551	5,551
Total Close	36,641	12,867	7,588
Distant Positive	16,347	4,389	985
Distant Negative	5,697	2,571	2,571
Total Distant	22,044	6,960	3,556
Total Close + Distant	58,685	19,827	11,144

Table 6.5: Rating Influence on Reading and Rating in Goodreads dataset

Category	**Total**	**Notable**
Close Effected to Agree	6,972	3,925
Close Effected to Read	5,199	1,763
Close Effected to Rate	24,470	7,179
Total Close	36,641	12,867
Distant Effected to Agree	2,692	1,783
Distant Effected to Read	2,274	607
Distant Effected to Rate	17,078	4,570
Total Distant	22,044	6,960

what to read, and dislike things their peers dislike more readily.

Next we examined the forms taken by the effect some actors had on others. Table 6.5 breaks down both close and distant instances of effect into three distinct categories. Firstly we see “Effected to Agree”, which describes those instances where an effected party rates an item in agreement with a socially-connected actor’s rating. Next “Effected to Read” measures the number of instances where an effected party, after an actor rates an item, marks that item as “to-read”. Lastly “Effected to Rate” is the effected party rating an item after an actor has marked it “to-read”. Unsurprisingly all categories are most frequently seen with close friends as activity like rating or marking “to-read” are visible to other users as part of the social stream of the Goodreads site. Instances of distant influence are also significant, we see many users rating items after an actor marked them “to-read”, indicating that actor could be the origin of a social chain that reminded their friend to rate an item.

Table 6.6: RMSE Accuracy of Socially-Aware Recommender Algorithm

Test Percent	Control	Close	Distant	Combined
20%	1.6741	1.6486	1.6737	1.6469
40%	2.4878	2.4530	2.4891	2.4506
60%	3.1409	3.1069	3.1412	3.1038
80%	3.7367	3.7122	3.7361	3.7097

Given that 30% of close and 18% of distant social relationships (as seen in the total influence in Table 6.4) resulted in an effected score from the average, we were eager to explore the effect these relationships would have on collaborative filtering recommendation. In order to measure the performance of the approach using three different signals (close ties, distant ties, and a combination of both) we examine the scores of RMSE, AUC (Area under Curve) and Recall. RMSE and AUC are designed to measure accuracy and performance, while Recall measures the capacity of an algorithm to retrieve relevant results.

Our tests looked at the entire recommendation list for each user in the system. We found, as detailed in Table 6.6, that close socially-tied instances of effect lowered the RMSE. Lower RMSE values correspond to improved accuracy which is notable across all tests, at all test percentage cutoffs. The distant information proved less useful with no notable change, either positive or negative, to the system accuracy. Combining the two signals performs roughly the same as using only the close signal in our tests, and this slight improvement is not notable in RMSE.

Further noteworthy are the recall figures in Table 6.7, where again the close signal outperforms the others, returning more relevant results. Close signal recommendation with 20% of the corpus reserved for testing generated 12,927 recommendations, as opposed to 9,473 in the control. This result shows more recommended items and greater coverage using social sources, with distant again not having a notable impact in our experimental setup. The AUC values are higher for close ties (Table 6.8), with higher areas corresponding to better performance.

Table 6.7: Recall of Socially-Aware Recommender Algorithm

Test Percent	Control	Close	Distant	Combined
20%	0.0074	0.0094	0.0079	0.0094
40%	0.0064	0.0079	0.0064	0.0080
60%	0.0108	0.0144	0.0108	0.0145
80%	0.0130	0.0202	0.0130	0.0204

Table 6.8: Area Under Curve (ROC) of Socially-Aware Recommender Algorithm

Test Percent	Control	Close	Distant	Combined
20%	0.3854	0.4019	0.3858	0.4024
40%	0.3679	0.3859	0.3666	0.3870
60%	0.3624	0.3671	0.3641	0.3698
80%	0.4706	0.4595	0.4698	0.4644

Table 6.9: Comparison of Socially-Aware Recommender tests (20% test)

	P@5	P@10	PRECISION	RECALL	RMSE	AUC
Control	0.0032	0.0033	0.1246	0.0074	1.6740	0.3854
Close	0.0025	0.0027	0.1453	0.0094	1.6486	0.4019
Distant	0.0035	0.0035	0.1246	0.0079	1.6737	0.3858
Combined	0.0025	0.0026	0.1464	0.0094	1.6469	0.4024

Lastly we present Table 6.9, which compares the results of numerous metrics for each approach. In addition to the discussed RMSE, AUC and Recall metrics here we show that the social context approach has an improved overall Precision (proportion of relevant to non-relevant items found) at the cost of lower Precision in the top 5 and 10 (P@5 and P@10 respectively). This suggests social context is a useful signal in situations such as conversational recommendation, where top-N recommendation (a flat list) is not the priority. Interestingly the "Distant" approach outperforms even the control for P@5 and P@10, perhaps because it focuses on providing more weighting in sparsely connected groups, leading to better information for top-N recommendation.

6.3.4 Discussion

As noted by Asch "...social influences shape every person's practices, judgments and beliefs ... a truism to which anyone will readily assent." (Asch (1955)), and here we have proof of these influences working in direct inter-personal relationships to form trends in opinions. The trail of influence outlined by socially-connected individuals highlights that social influence is a detectable signal that could be leveraged for recommendation. That a detectable connection exists at all is of great interest as it shows a relationship between two people that not only takes the form of one person being informed by another, but also of a person having the power to affect another's enjoyment of something before they evaluate it. From this we can see that a person can, through expressing their like or dislike for an item, effectively spread their opinion to others, or at least influence how others perceive the item. This work may be useful in identifying not only trends but also trend-setting people.

This finding proved that social context, the idea that a person's place in a social network could influence their ratings and enjoyment of an item, can be detected and could be useful as a source of information for recommender system modelling. It is interesting to note that there were many more instances of negative post-experience (i.e. after an actual rating from both parties) effect in the dataset. This could be the manifestation of the customer service adage that "a happy customer tells one friend, an unhappy customer tells everybody.", in that while these days equal numbers see both negative and positive reviews, it is the negative opinions that tend to have a greater effect.

What we have been calling influence here can be seen to be a confluence of concurrent signals all coming from social interactions. If a user shares a rating where all their friends will see, we have shown that in some instances their friends will agree with them in a detectable way. We have also shown that still others will be interested enough in their friend's shared item that they themselves will mark it for future examining. More will be reminded on seeing their friend sharing an intent to read a book that they have read it, and rate the book there and then. Our approach of using a local average to signify the combination of opinions turns out to be indicative of this behaviour from the outset, with either party choosing a "to read" status resulting in a local average that is diminished leading to a lower weight, effectively they haven't been influenced in opinion but rather in what item to read.

It is interesting to note that although distant influences are apparent they do not manifest themselves when used alone in the recommender tests, perhaps because the parties are too distant and have differing interests. This means that with our current experimental setup, using 5-fold validation at our test percentages with weights based on the difference from the local mean, did not find an impact of socially-distant people, though these users constitute a detectable signal. This means that user-based collaborative filtering using Pearson correlation as a distance measure may not benefit from this signal, though it is present. A conversational recommender could make further use of this social signal by adopting desired influencers and examining the disparity (if any) between real influence and desired influence.

Authors were not leaders of popular opinion among connections in our experiments to explore the ways their presence may prove beneficial to recommendation. In fact none of the influencers in our initial experiment were authors, which we wished to investigate further. As social annotation begins to be examined in search with limited success (for example Muralidharan *et al.* (2012)), we were interested in the potential of social annotations in recommendation, in for example, a conversational system that would ask about what a person thought of their friends' influence, and influence detected.

The Goodreads dataset offered a pre-annotated vector for this sort of social impact study, that of authors. This afforded us a chance to examine methods by which a person's favourite author or expert whose opinions they believe they agree with could be taken into account. Having looked at both observed social influence and its suitability as a recommender data source we wished to further examine another avenue afforded us by this dataset.

6.4 Weighting Social influences

So far we have shown multiple effects of a number of social relationships between people, but in this section we focus on a different kind of relationship. Here we look at improving the core recommendation algorithm that in Chapter 3 we showed works well with conversation. People have implicit relationships with so-called 'content-producers' in that we consume the content we like. These relationships are developing in the post-Web 2.0 world, as evidenced by the Goodreads dataset which annotates authors as a special subset of users. We considered how peoples' recommendations many benefit from their perceived or detected relationship to the authors of the books. This section also endeavours to detect whether authors could be considered experts, whether they could be used in expert-based recommendation, as detailed in Amatriain *et al.* (2009). Such work as Heath (2008) has found that people choose sources for their expertise, experience and affinity to the subjects they talk about, three characteristics that an author usually has. Our previous work on social trails made use of authors as if they were regular users, but here we examined authors as possible experts on the subjects people are rating: books. There are 2,747 authors in our Goodreads corpus. This type of user division is not unique to Goodreads. For example many Twitter users are “verified” recognised celebrities who frequently endorse things they like in a manner similar to rating.

6.4.1 Approach

We looked at how coherent an author’s interests are in terms of books they review. This is of interest as an author could be considered by other users to have good general knowledge of books. Users may believe they know what makes a good book or at least a good book in their specific genre. We consider whether a person is influenced by authors based not on the social reach of the author but on the relationship between the person and the author. In doing this we are interested in the influence or popularity of authors controlling for their followers on the site, if not the number of people rating their work.

Beyond this we examine another aspect of how expert users such as authors may use social networking sites to review things. We were interested in how these experts share opinions in a domain where the public can see them comment on their colleagues’ work. We wished to see if experts affected a review site in the same way celebrities can have a huge impact, e.g. causing controversy by making negative comments about others or drawing huge attention to websites by linking to them, on Twitter. This might also be used as a method of detecting malicious or under-handed uses of recommenders to highlight an agenda or strategy.

Further we examined how experts used this perceived social power both in their average review scores and what they review. We examined what content experts passed comment on or rated, looking at convergent authors, those who mostly rate items with similar genres to their own writing, and divergent authors who rate a range of things. We then compare the effects of these expert users on the users they effect, in total and in each category.

We were interested in how the range of author opinions might affect user reaction to them. Convergent authors, those who review mostly books of the same genre as they write, might appear knowledgable on the genre or solely interested in expressing opinions of their colleagues' work. Similarly divergent authors might be viewed as either just another person or someone with authority above the average. We examined this divergence using the tags associated with an author's books and the books they rated, focussing on overlap.

Having explored this measure we then considered two methods by which authors could potentially socially influence others above the measures of regular users examined in the previous section. We considered that users without direct connections to an author might believe they have an interest in the opinion of, or connection with, authors whose work they read. Further to this we looked at the special status of authors and explored artificially weighting the opinion of the author most similar to the user to see the effect of giving experts more power within the crowd.

In order to test if author weighting improves recommendation we formulated two methods to weight recommendations based on what authors like. We began with our control, a user-based collaborative filtering algorithm using Pearson correlation as a measure of similarity. For each tested approach we performed five-fold cross-validation. Our first approach was to treat authors as users whose opinions matter more to others with similar taste. We therefore examined each user and weighted a single author who was most similar to that user, shown in Algorithm 4. Weighting worked the same as in our prior experiments, adding to similarity rather than replacing.

Next we looked at a different strategy, reasoning that authors a person had experience with, in this case had read the work of, might have opinions of interest

Algorithm 4 Author Similar Weight Generation Algorithm

$weightval \leftarrow weighttobeaddedpercommonality$
for all $User$ **do** ▷ Weight the author the user is most similar to
    $Weights[MostSimilarAuthor][User] \leftarrow weightval$
end for
return $Weights$

to them. Algorithm 5 is not limited to Goodreads, as it could apply to celebrities a user on Twitter follows, has had interactions with, or similarly detects marks of actual interest. In our experiments "weightval" was set to 0.2 to push similarity closer at each similar social connection.

Algorithm 5 Author Read Weight Generation Algorithm

$weightval \leftarrow weighttobeaddedpercommonality$
for all $User$ **do**
    for all $BooksReadByUser$ **do** ▷ Weight all the authors the user has read
        $Weights[AuthorOfBook][User] \leftarrow Weights[AuthorOfBook][User] + weightval$
    end for
end for
return $Weights$

Since the purpose of this analysis is to examine the ability of experts to improve recommendation we also looked at a potential classification of their expertise in how the authors rated items. Since books in the Goodreads dataset are tagged with genres we were easily able to look at the work an author wrote and compare it with the work they rated through the genres each item is tagged with. Genres for books in the Goodreads collection appear to be at least partly folksonomic in nature, i.e. there are a limited set of genres that are officially given to items based on which tags users apply to that item in their own collection. Users "place books on shelves", tagging them with whatever shelf name they wish, which is frequently a genre or sub-genre ("high-fantasy"), but could also be "to-read","currently-reading" or "didn't finish". Through some form of curation only commonly used genre tags are applied to the items. We can then compare and contrast the works of different experts. We did this to see if the authors' interests, and therefore possible usefulness, lay in a broad range or narrow single area.

We termed these types of interest *divergent* and *convergent*, seeing as neither could be said to be obviously preferable. We defined convergence as above 50% of the rated item tags are the same as tags present in the authors' own works. These tags ranged from genre labels to loose conceptual tags, meaning some tags had greater meaning than others. Using this definition there were 152 convergent and 2,595 divergent authors in the collection.

Following our experiment with weighting the opinions of authors based on similarity or user experience of them, we examined integrating the convergence metric we had defined. We were interested in seeing if more positive impact was felt from experts who specialised in a sub-field of their area or experts who generalised. To this end we looked at weighting convergent specialists and divergent generalists separately to see if either group offered greater clear benefits. For this we revised our "Authors Read" (Algorithm 5) and "Authors Similar" (Algorithm 4) weighting algorithms to separately weight convergent and divergent authors, and we compared the results.

6.4.2 Results

We measured the performance of our approaches using standard metrics, Root Mean Square Error(RMSE), Precision, P@5, P@10, Recall and AUC, over the entire user set (including authors to see author to author influence). Unless otherwise stated (P@5, P@10) figures were computed using the entire recommendation list for each user. Our first set of results are a full comparison between the "Authors Read" (AR), "Authors Similar" (AS) and the Control, a user-based collaborative filtering algorithm using Pearson correlation to determine user similarity, the same algorithm we modified with both weighting strategies. For each test we withheld a percentage

Table 6.10: RMSE values of Social-Role-Aware Recommender Algorithm

Test Percent	Control	AR	AS
20%	1.6741	1.6782	1.6744
40%	2.4878	2.4883	2.4894
60%	3.1409	3.1409	3.1415
80%	3.7367	3.7367	3.7376

Table 6.11: Area Under Curve (ROC) values of Social-Role-Aware Recommender Algorithm

Test Percent	Control	AR	AS
20%	0.3854	0.3831	0.3959
40%	0.3679	0.3571	0.3718
60%	0.3624	0.3762	0.3682
80%	0.4706	0.4718	0.4581

of the ratings in the collection to see how well each approach could predict them, as indicated in the tables.

Table 6.10 shows our RMSE comparison. It is clear from these numbers that in our experimental setup neither AR nor AS approaches offer either significant benefit or disadvantage over the control. Equally, the Area Under Curve measurements (Table 6.11) show performance not measurably different from the baseline in our tests.

In our measurement of Precision, as shown in Table 6.12, both AR and AS methods were notably worse than the baseline, while Recall (Table 6.15) proved no better. This indicates that AR and AS find less relevant results within the collection. P@5 and P@10 results (Tables 6.13 and 6.14 respectively) also show no significant advantages, except for a slight improvement at high test percentages, indicating that in rating-sparse environments, AR and AS methods may offer improved Top-N recommendation in cold-start situations where little about the user is known.

We now look at the results of weighting based on whether the authors are convergent or divergent in their interests, integrating these into both AR and AS methods. Again we tested across RMSE, P@5, P@10, Precision and Recall metrics for AR

Table 6.12: Precision of Social-Role-Aware Recommender Algorithm

Test Percent	Control	AR	AS
20%	0.385430155	0.12298	0.12333
40%	0.367860826	0.08565	0.08580
60%	0.362358916	0.05319	0.05234
80%	0.47062143	0.00927	0.00933

Table 6.13: P@5 of Social-Role-Aware Recommender Algorithm

Test Percent	Control	AR	AS
20%	0.00325	0.00305	0.00339
40%	0.01151	0.01237	0.01263
60%	0.04363	0.04156	0.04157
80%	0.17297	0.18395	0.18436

Table 6.14: P@10 of Social-Role-Aware Recommender Algorithm

Test Percent	Control	AR	AS
20%	0.00333	0.00337	0.00324
40%	0.01223	0.01295	0.01311
60%	0.04482	0.04277	0.04365
80%	0.18213	0.19627	0.19720

Table 6.15: Recall of Social-Role-Aware Recommender Algorithm

Test Percent	Control	AR	AS
20%	0.00742	0.00258	0.00259
40%	0.0064	0.00622	0.00627
60%	0.01078	0.01114	0.01077
80%	0.01299	0.01273	0.01283

with Convergent (ARC), AR with Divergent (ARD), AS with Convergent (ASC) and AS with Divergent (ASD) and again the results showed little positive or negative impact. Appendices 6.16, 6.17, I, II, III, IV, V, VI, VII and VIII show our findings in detail.

Table 6.16: Convergent vs Divergent Authors Read (RMSE)

Test Percent	**Control**	**AR**	**ARC**	**ARD**
20%	1.67405	1.67821	1.67664	1.67398
40%	2.48782	2.48834	2.48751	2.48938
60%	3.1409	3.14089	3.14011	3.14288
80%	3.73673	3.73667	3.73559	3.73639

Table 6.17: Convergent vs Divergent Authors Similar (RMSE)

Test Percent	**AS**	**ASC**	**ASD**
20%	1.67436	1.67552	1.67691
40%	2.48938	2.48677	2.48966
60%	3.14145	3.14174	3.13965
80%	3.73758	3.73704	3.73558

6.4.3 Discussion

Having performed a full assortment of tests to assess the usefulness of experts as they are detected within our Goodreads dataset we now discuss the results. Which metrics to use in order to perform as objective an analysis as possible is still an active area of discussion (Felfernig *et al.* (2011)), but from what we can see here through common measures, at our given experimental settings, nothing conclusive was found for either read or similar authors as influencers. Some minor improvements may be obtainable in sparse rating environments, but otherwise there were no measurable improvements or losses. A possible reason for this is that Goodreads has a separate “fan” category as distinct from a friend, not examined in our dataset due to its specific application to Goodreads and therefore not easily generalised to other datasets. We wished to examine friend relationships rather than fans, which are semantically different.

The average user had 194 friends on Goodreads, while the author average was 48. This reduces their immediate social graph, which in our prior section was shown to have the highest impact, possibly limiting their ability to affect widespread opinion. Since little influence is seen using this algorithm a different method, either algorithmic or experimental might need to be employed in order to better use authors as experts, if one exists that is not dependent on the design features of Goodreads. Further exploration of the experts' interests found no improvement. We did no semantic analysis or distance measure within tags in the collection, resulting in labelling "fantasy" and "high-fantasy" as just as different as "action" and "romance". This was enough for our purposes to test the concept but the results might be different with a different approach.

6.5 Comparison to Related Work

6.5.1 Social Trail

Our work investigating social connections is similar to recent work reported by Bourke *et al.* (2011), in which the authors studied social connection and its ability to generate recommendations. In that paper the authors examined various neighbourhood selection strategies as the primary method of recommendation, where we weight based on perceived impact of neighbour opinions. We also look at incorporating a person's social history, in a similar way to browsing history Matthijs and Radlinski (2011), into the weighting process. Other work, Liu and Lee (2010), has looked at combining social connections with collaborative filtering, but we here compute the value of each relationship based on how well the influencer predicted the influenced's ratings in any common items they rated first.

Much work has been done in the area of trust for recommenders (such as by O'Donovan and Smyth (2005)), including in social networks (Golbeck and Hendler

(2006)). In some ways our work is similar to trust measures, in that it looks at the impact of one user on others, but there are distinct differences. Our approach is novel because it is interested in user effect on users, not the system as a whole, and imposes temporal order on any connections inferred (which can only happen through social ties). This is in order to trace the origin of a user's difference of opinion or to spot trends, as well as to identify users who are influencers or mavens, rather than simply useful for recommendation. In situations where the set of commonly rated items between people is sparse, correlation-based approaches can falter, and this is where trust features can help. Work has been done Massa and Avesani (2004) (later developed and evaluated Massa and Bhattacharjee (2004)) to explore how even simple trust relationships can increase coverage. We see in the recall numbers that our approach also improves coverage, as there were a much larger number of items recommended.

Most frequently, social network recommender systems use how much a person trusts their friends, or the opinion of their community to recommend items, a sort of *community pulse* (Terveen and Hill (2001)). Here we analyse the direct influence, or how well one party (either distantly or closely connected) predicts another's rating, and examine the use of this information source to improve recommendation by weighting.

Our concept of social trail grows from social recommendation work that builds recommendations by scraping real-time social sources such as Twitter (Esparza *et al.* (2012)). Previously much work has been done on detecting trustworthiness in social recommendation (for example by Golbeck (2005)), that is how much one person should trust another to whom they are not connected. Here we are not concerned with trust but accounting for already apparent influence that impacts the user's opinion, thereby altering the ideal recommendation.

6.5.2 Expert Authority

It has been shown that a small number of experts can improve recommendation (Amatriain *et al.* (2009)). More recently in trend identification and recommendation work (Sha *et al.* (2012)) has been done to capture the wisdom of the few people whose opinions hold real affecting weight, while other work has been done to examine social context (Ma *et al.* (2011)). In the Goodreads dataset we had an annotated corpus of people, a portion of whom were authors. These authors have expertise around, experience of and affinity for books, three key factors in source selection (Heath (2008)). They also represent an authority rather than simply a trusted source, as studied in Passos *et al.* (2010). We looked at not selecting sources based on this knowledge but weighting the expressed opinions of authors. Others (Kazienko *et al.* (2011)) have looked at semantically different relationships but here we looked at different social roles within the dataset. This could equally be applied to Twitter (through either "verified" account status, follower numbers or semantic analysis) in order to apply our approach to another dataset; the Goodreads author/user relationship has analogous relationships across the social web.

In other work (He and Chu (2011)) trust issues between the user and the system that relate directly to this work are described, in that the authors identify "Misleading by Friends with Unreliable Knowledge" and "Shilling Attacks from Malicious Users" as issues in social recommenders. This has lead to work looking at reputation, including research by McNally *et al.* (2010). Here we investigate what might be considered social trust in expert usefulness, where experts are not necessarily going to provide good information without an ulterior motive (we know for example that celebrities are paid to send messages on Twitter promoting products, which may be seen as introducing bias). We did this through contrasting the tags the experts are considered to have expertise of and the ones they rate, drawing inspiration from other applications of tags in folksonomic domains (where users create and manage tags) Gemmell *et al.* (2009). One motivation for our work was to see if convergent authors, who mostly rated within their genres, would reduce recommendation accuracy because they were rating the work of colleagues for their own gain (which could be termed shilling or misleading other observers). This requires further investigation but is outside the scope of this thesis.

6.6 Chapter Conclusions and Answer to Research Question

RQ 3 Can social relationships inferred from contextual cues prove useful in improving recommendation accuracy?

In this chapter we show that social context, as derived from the temporally organised commonality between socially connected users can prove useful in certain circumstances to improve recommendation. We showed connected people offer a way to discover new items, but also that after experiencing an item a friend's opinion could influence evaluation of that item at rating time. We then took that knowledge and weighted a recommender system to show that the information from closely connected friends can help improve recommendation, while we did not find a way to leverage distant relationships. We also showed that as the social context approach improves Recall but not P@5 or P@10 scores it is well suited to conversational recommendation rather than Top-N recommenders. If we had this knowledge *a priori*, through annotation or conversation related to social connections the analysis we perform here could be used to improve recommendations offered to users. We then looked at different kinds of social roles within the collection, examining people known as "authors" within the book recommendation domain to see if they improved the results of people similar to the author or that read the author's work. Here we found no noticeable difference from the baseline. Examining these factors we showed that not only simple social relationships but *who has rated an item before a person*, though not necessarily who that rater is seen to be in terms of expertise, can be detected and used to improve recommendation accuracy.

Chapter 7

Context

7.1 Context and Recommendation

In the previous chapter we looked at social context and how it is, and can be used in recommendation. In this chapter we turn our attention to an examination of traditional context and how it is used in recommendation. Context can be defined as information that modifies a person's understanding of their current situation, or affects their current choices. Context-enriched services are becoming more and more valuable as people now adopt new habits in their usage of context-aware technologies, for example allowing mobile activities such as checking-in at new locations.

Context has already been explored in recommendation (Adomavicius and Tuzhilin (2011a)), but here we look at sensed context in a different way. We are interested in individual views on context as it pertains to recommendation. The usual approach to context has either been to take every sensor available or to design the context sensors used around the task.

Before we go further, we should explain the term contextuality; textuality is the attributes that distinguish the communicative content under analysis as an object of study, contextuality is the contextual sensors that are optimal for a particular user within the system. So for example one person might make choices by taking location into consideration, while another might not feel location has any bearing on the situation or choices to be made. We are interested in attempting to detect what type of user a person is and predicting what contextual attributes will best mirror their own decision-making process in order to better offer item suggestions or recommendations.

Our research question asks if this user-level (rather than task-level) unique context set for a given user at a given time and in a given situation, can be seen in a system with broad contextual sensing. In order to examine this first we need to know does contextual recommendation benefit from picking and choosing its sources to begin with? Recent work by Baltrunas *et al.* (2011) shows that it does, as many contributing contextual factors are frequently unnecessary for a task. Next we need to know what do users want to share as context? This will inform what is acceptable to use as context in later tests and speak to how people make use of contextual recommender technology. Our first experiment in this section deals with this.

Finally, having established the degree to which users are comfortable sharing context, can we determine a context selection strategy from that context alone? We wish to find out if we are able to use context to choose the best set of contexts to use for a person to offer high-quality recommendations. Having looked at each person's best contextual fit we were then able to comment on how this affects the system as a whole, if any trends were visible that could be used for everyone.

7.2 Shared or Sensed Context in Conversational Recommendation

The idea of somehow capturing and using a user's context as s/he uses some computer system spans multiple disciplines, including psychology, philosophy, anthropology as well as the technical aspects in engineering and computer science. Generally the term context-awareness denotes the ability to ambiently capture and make use of the user's context without interfering with the task the user is trying to accomplish (Dey (2001)). Each field that has explored context tends to take a different approach to the subject, with anthropologists and sociologists conducting ethnographic studies (for example, work by Goodwin and Duranti (1992)) and a great deal of computer science and engineering work concerned with the methodology of collecting and using directly sensed data from the subject.

The importance of knowing context in any kind of user interaction cannot be overstated, as it is the means by which users and systems come to a mutual understanding. Derrida, whose field of deconstruction probes the context of works, said "There is nothing outside the text" (Derrida (1976)), which he later explained as "There is nothing outside context". From a HCI perspective this can be seen as foreshadowing the usefulness of contextual data in driving the over-arching narrative of interaction within a system.

Context-awareness is a key requirement of human-centric computing systems, allowing them to adapt and to form meaningful interactions by accounting for the user's current needs, task, environment, etc. Yet there exists an issue; purely sensed context needs a great deal of data to infer patterns of usage and meaning, for example GPS coordinates could tell that a user visited a shop twice, which could either mean they are a frequent customer or they bought something that was faulty and had to be returned, meanings that imply vastly different levels of customer satisfaction for example.

Barkhuus and Dey (2003) explored and defined three levels of user interactivity related to context-awareness: personalisation, passive context-awareness, and active context-awareness. Personalisation makes use of user settings, whereas context-aware applications make more dynamic use of context or sensor information. Active context-aware systems automatically make context-based changes, which Barkhuus and Dey found through evaluation to be preferable to passively offering the option

Table 7.1: Survey questions

Question	Possible Answers
What are you here for ?	- just browsing - looking to buy - sharing my opinion
Are you in a group ?	- just me - me and a friend - part of a couple - party or big group
Where are you ?	nowhere important - point-of-purchase - researching

to change. Our work explores the collection of this data.

7.2.1 Approach

Our experiment in context-gathering made use of a recommender application to help users find movies that might be of interest to them, a system we described in Section 3.3.2. During an on-line evaluation of our system, users logged into the website to use the recommendation system. The users participated in an average of 9.1 sessions within the system, each time beginning by answering a brief survey. The survey asked them the purpose of the recommendation. We asked three multiple-choice questions of users to put their next interactions in context within the system. These questions were tailored to the task in order to greater understand the users' needs and actions and these are shown in Table 7.1. They correspond to a general changing of intent, as if the user was donning a different profile depending on answers they gave (indeed this is how we envision this approach generalising). Instructions to the users explained the rationale behind them. Importantly, the questions demonstrate the intent behind a context, i.e. "I am here to browse", distinct from the sensed details of "I am in a shop" or even "I am in the large music shop on Y street in X city". This was in order to supplement any automatically-sensed data and provide a more conceptually accurate context.

Table 7.2: Context statistics

247	users
614	sessions
4.1	average context entries per person
149	entries of sensed context
30	different operating system/browser combinations
864	entries of surveyed context

At the start of each session we also recorded location as available (using HTML 5, which gave GPS for mobile users or approximations for desktop users), operating system used on the device, browser and IP address. Depending on the browser security settings, a user could choose to not share their sensed data with the system, although in their instructions we warned of this and asked them to share the information.

7.2.2 Results

The summary data is shown in Table 7.2. Over the 247 users the mean collected different sensed data was 3.5 (indicating a relatively similar purpose over the 9.2 interactions). This could point to methods of surveyed context as user profile in a shop for example. Importantly it can be seen that over the sessions only 149 times did the users allow sensed context to be gathered, even with the knowledge that it was wanted as part of the test.

From the figures in Table 7.2 we see that users more readily answered the survey than shared sensed data. In less than 25% of cases the user choose to share sensed data, indicating an issue of trust with the system. The survey generated a large number of responses as it was a key step in the system. Almost 30% of the collected survey answers are different from the default, indicating the need for good defaults that make sense. In our case we allowed for the possibility that the user placed no special value on their current context.

After the online evaluation we asked 34 of the users about the system. 28 said they would use it again, showing a general acceptance for this sort of mechanism for capturing context via dialogue. Our method of conceptual context shows potential for framing a single use of a recommender system as part of a larger narrative, for example "This user likes vastly different films when they are browsing with their partner". By focusing the user on interacting with the system they are comfortable sharing beneficial information that they are unwilling to share through direct sensor activity, and have some understanding of how context is viewed by the system. User trust in context-gathering is an area that needs to be further explored.

7.2.3 Discussion

When users respond to recommendations with ratings or other straightforward interactions such as "likes" this can represent a missed opportunity to capture what could be a deep personal expression of an opinion on a recommended item. From the preliminary work that we have reported we found that giving users a method by which we can provide a frame of reference for these opinions and allowing a richer kind of user feedback appears to be a positive thing, as long as the system is careful not to impose meaningful context when none is perceived by the user.

Our focus in this thesis is understanding ways in which context can play a part in each persons' recommendation experience, and how differing views of context can be accounted for. We established here that people make use of sensed and surveyed context, which leads to our next question; can we determine a context selection strategy from the context alone?

7.3 Contextuality - Context Sets and their Usefulness

Having looked at how people feel about sensed and surveyed context we wished to explore how useful people find context. Until recently it has been assumed that in contextual recommendation all forms of context available should be used for any task. Recent work has shown that some contextual information is irrelevant for some tasks, here we have investigated whether individual users have preferences for optimal context set for them within a system.

As we have shown in Section 5.3 users can be in a position to reject otherwise good suggestions, so any contextual features that could account for or alert us to this fact should be of interest. Problematic though is the fact that memory-based recommender systems focus on forming groups of users from what is known about them, essentially stereotyping people, and the more we know in the form of contextual data the hard it is to decide how to form groups. Contextual data might be important by design for the given task, or different information might be important to different people.

7.3.1 Approach

Our experiment is designed to highlight the contexts people are interested in when following a user on Twitter[28]. Twitter is a social network micro-blogging site that allows a user, under a screen name, compose 140 character messages for people following them to read. Users have followers and friends who they follow to see updates (called tweets) from. Other features like marking a tweet as "favourited", putting users in lists and "retweeting" (sending a message from someone else to all your followers) also exist. Many of these user-generated micro-blog streams are publicly available.

[28] `http://twitter.com`

We collected a dataset of tweets from publicly-accessible twitter users, using the "firehose" Twitter API. We gathered 251,807 tweets from 7,390 unique Twitter users within the Dublin area. We restricted our collection of tweets to one area in order to control for timezone, as we examined the times people tweeted.

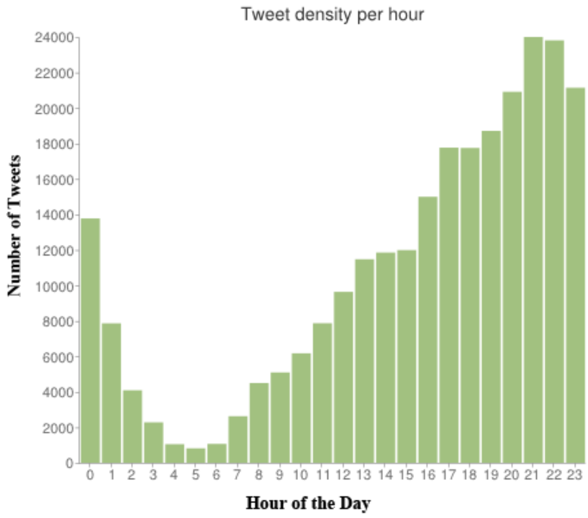

Figure 7.1: Tweet density over time, from public Dublin-based Twitter users over time

Twitter provides a wealth of data with each tweet. We took 61 features (shown in Appendix IX) used to describe users of the service in their tweets. In keeping with Section 7.2 of this chapter we included in our contextual features anything that told us about the user that was freely provided. This ranged from those that were sensed (for example their location details) to those that were readily shared with the world (their Twitter biography), all accounting for the context of how that user presents themselves to others. We took 37 features made available in the tweets (such as the source; which client sent the tweet) or otherwise computable from the features available. Where we knew the feature would be unique (such as the screenname or real name of a person) we computed features that would make these fields comparable (detailed in Table 7.3). In addition to these 37 features we had 24 features to characterise how many times the user tweets in each hour of the day. For the purposes of using machine learning we categorised each of the text features

Table 7.3: Descriptions of Dynamically Generated Features

Dynamically Generated Feature	**Description**
Capital letters in screenname	Number of capital letters in user's screen nickname
Capital letters in name	Number of capital letters in user's actual name
Description length	Number of characters in the user's biographical description
Name length	Number of characters in the user's name
Screen name length	Number of characters in the user's screen name
Screen name is real name	Is the user's screen name equivalent to their real name?

with a number, Table 7.4 details the number of categories generated for each of the text features. Other, numerical features, were used that did not need categorisation. These are listed in Appendix IX. This preprocessing gave us a list of 7,390 users as described by the context they present to the world, that they tweet only at certain times, or are popular or unpopular (based on follower count or similar metrics).

For the purposes of our experiment we were interested in who each user in the collection followed, and what contextual data might have influenced that decision. We gathered each person in the collection's complete friends list. This allowed us to highlight which people in the collection followed each other. We were then able to generate for each person, a list of every other user in the collection as described by their contextual features, annotated with whether or not that person follows them. This preparation left us with the data formatted for the tests we wished to perform.

We first looked at the importance of each feature as a means of discriminating within the set for each user. F-score is a simple technique which measures the discrimination of two sets of real numbers, as described by wei Chen and jen Lin (2005). The larger the F-score is, the more likely this feature is to be more discriminative. It is important to note that if one user exclusively follows people with low tweet counts and another exclusively follows people with high tweet counts then both will

Table 7.4: Text features and the number of categories for each

Feature	Number of Categories
Geotype	1
Language	11
Location	2552
Place full name	38
place id	38
place name	35
place type	3
place URL	35
profile back colour	1089
profile sidebar colour	1116
profile sidebar fill colour	1180
profile text colour	1021
source	101
timezone	75

have high F-scores, as "number of tweets" is a very discriminative feature for both. We calculated the importance of each feature for every user, then averaged them over all users. This will form an integral part of the feature selection we perform later. For each person within the set we computed their individual F-scores based on who they followed.

Having examined F-scores we then proceed to perform feature selection for a random group of 530 users from the collection. We wished to see what influenced whether one person followed another, in order to potentially offer better contextual recommendation. We used this data to build an SVM per person to model their individual interests, using libSVM (Chang and Lin (2011)[29]). Training used the entire list of users with the 61 features and whether or not this user follows them. We categorised all of the text-based features into numerical format in order to be compatible with the SVM training. We used the feature selection tool provided with libSVM[30] to rank the important features in the dataset. After we ran feature selection on each user, we took the minimum number of features necessary to accurately

[29] http://www.csie.ntu.edu.tw/~cjlin/libsvm/
[30] http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/

Table 7.5: The top average important features in deciding whether a user follows another

Feature	**Mean Strength**	**Std Dev**
Follower count	0.01147	0.0689
Listed count	0.00673	0.0236
Friends count	0.00260	0.0402
Favourites count	0.00243	0.0077
Statuses count	0.00147	0.0037
Posts during 16:00	0.00093	0.0070
Posts during 19:00	0.00068	0.0048
Posts during 17:00	0.00067	0.0039
Posts during 21:00	0.00065	0.0038
Posts during 20:00	0.00065	0.0036
Posts during 22:00	0.00065	0.0035

produce the same results in order to arrive at our final analysis.

7.3.2 Results

Examining the F-scores we averaged the scores for each feature over all 530 users and found, as detailed in Table 7.5, that each of the most important features has a high standard deviation, indicating importance of features is very personal to each user. We do see that on average, follower count is clearly the most discriminating feature.

Having looked at the most discriminating features available, we trained 530 SVMs, one for each user. These SVMs were trained on the prepared list of each users' contextual representation, annotated by whether or not the user the SVM is modelling follows them. In all but three cases, users' following habits were indicated by only three features. The three special cases include one user who required 13 features and two that maintain their highest accuracy with six features. Table 7.6 shows an aggregated count of features as they appear across each user's feature selection set. This corresponds to how the user evaluates who they follow. Follower count and Listed count, both highly discriminating features overall, top the list, but

Table 7.6: The most selected features by SVMs trained on individual users

Feature	**Number of Users For Whom Feature was Selected**
Follower count	185
Listed count	174
Profile background is tiled	90
Description length	81
Statuses count	64
Screen name is real name	63
Geotype	58
Favourites count	53
Name length	51
Profile text colour	49
Friends count	47
Location	39
Capital letters in screenname	33
Capital letters in name	32
Profile sidebar border colour	30
Posts during 12:00	29
source	29
Profile background colour	27
Posts during 14:00	23
Place name	22
Posts during 7:00	21

other features that may not be as obvious, such as whether the profile background of a user is tiled, play a part in defining how one user sees another.

7.3.3 Discussion

We have shown here that there are distinct groups of users who use different sets of contexts. Depending on the user we can recommend the *context set* they should use, in order to improve contextual recommendation. This could conceivably lead to modelling users based on what criteria they use to evaluate the world, a "context-profile" that could accompany people in the cloud to be used by any service that recommends using context. This would easily generalise over contextually-relevant tasks, as nothing about our experiment was specific to Twitter, which we used for the availability of a range of context data.

We have highlighted in Twitter that follower count is a decisive metric for user interest in following. However it is only seen as important to 185 of the 530 people who were analysed, indicating that it would not improve recommendation for the majority of users. If there had been some solid consensus on which features to use this would be a valid method of using context to choose the context to use when recommending. For a user of Twitter this might mean that the contextual friend recommendation process would evolve, so they could be grouped with others who have similar context-requirements based on their actions and therefore only use the most discriminating contexts for their recommendations. If further investigation found this to be a wholly positive correlation (i.e. people always valued more followers) this could speak to the suitability of collaborative filtering for Twitter user recommendation, as sparse ratings (or less followers) would actually be indicative of a trend toward a less suitable recommendation.

Furthermore, it is interesting to note that while no contextual feature provides good coverage of the 530 users (i.e. no one feature could be used to predict accurately), sets of contexts do reoccur, opening up the possibility of using a recommender system to class users based on their behaviour and recommend a set of contexts that will most likely improve their recommendations.

7.4 Comparison to Related Work

7.4.1 Views of Context

Context has long been discussed as a useful data source in many computer tasks as discussed by Lieberman and Selker (2000), and contextual recommendation has a rich background of related work (Adomavicius and Tuzhilin (2011a); Dey (2001)), making use of sensed data such as location or time to improve the quality of the items recommended. While the distinction between "active" and "passive" modes of context use is made clear in Barkhuus and Dey (2003), here we explored "transparent" and "opaque" modes of context *collection*. Gathering context from sensors transparently and ambiently so the user does not even have to be made aware of the collection process and where it does not interfere with the user's task, is the current standard (see work by Athanasopoulos *et al.* (2008)). In an attempt to aid the definition of semantic meaning around this context-sensing data, we built a system to test a method of querying the user prior to system interaction, opaquely gathering the reason behind the data gathered. Rather than trying to describe context in terms of a set of features associated with the type of device, location and date/time, we model context as a hidden process that at any time can be in one of a finite set of states that have a bearing on the user's behaviour, in a similar way to Anand *et al.* (2007).

People have a cultural understanding of context, both in complex constructs of language (as discussed by Goodwin and Duranti (1992)) and social situations, abstract concepts such as what is acceptable in public versus private (Warner (2005)). Since context is such an abstract concept, information that forms a context can be represented in various formats. Much work has been done in computer science to provide middleware (e.g. Athanasopoulos *et al.* (2008)) to fuse the multitude of contextual sources a system might need in order to be fully context-aware. Here we looked at giving the user a method to express the meaning of their own context along with contextual data collected, providing semantics at the point of collection, rather than after collecting enough data to determine if there are patterns. The idea of modelling for a more complex view of context is not new (Schmidt *et al.* (1998)), indeed it has been broached as a sensor fusion problem before, but here we find a possible benefit of users expecting interaction; they are willing to tell us about their perceived context.

7.4.2 Contextual Feature Importance

The place of features such as sensed context (then considered as part of a measure of performance) has been debated since before sensors became as sophisticated as they are currently (Newman and Newman (1997)). Here we showed it is possible to measure the performance of represented contexts such as place, time and online identity features for each user of a system.

It is well known that choice is affected by context (investigated by Yoon and Simonson (2008); Dhar *et al.* (2000)), which could be for a number of reasons, perhaps involving inconvenience (tying in with Connaway *et al.* (2011b)), in that context can be a barrier to making certain choices. As has been mentioned earlier in this chapter only some contextual features are relevant for any given decision within recommendation (Baltrunas *et al.* (2011)), and work done by Madani and DeCoste (2005) highlights that not all context impacts recommendation. Here we turned our attention to user-level contextual feature selection, finding that each user is indeed different in the features they considered. In the past designing for context has been styled as scenario oriented recommendation, in that recommenders are then only useful in the envisioned scenarios (Shen *et al.* (2007)).

Recent research by Wilson (1999a) has defined three major methods for incorporating context into recommendation algorithms. These three methods are pre-filtering, post-filtering and altering the user model. While none of these methods offer a clear enough advantage to abandon the others (Panniello *et al.* (2009)) none provide a method to determine which contextual factors are of primary importance dynamically. Recommender systems built to be "context-aware" such as discussed by Adomavicius and Tuzhilin (2011b) would further benefit from being "user-aware" in the choice of that context, as we have investigated here. Machine learning is not new in recommendation (Breese *et al.* (1998)), but here we apply it in a novel way. Previously contextual recommendation has used a single SVM to model context over all users (Oku *et al.* (2006)), here we train an SVM for each user to examine how each user benefits from each feature. We do this for much the same reason as Noulas *et al.* (2012) conducted their research into modelling context using random walks; the problem of an abundance of contextual data available to improve recommendation becoming available from a variety of sources. This work can be seen as an extension of work by Koren (2008) into latent factors in recommendation, but applied to the new area of contextual factors.

7.5 Conclusion and Answer to Research Question

Earlier in the thesis we set out our 4th research question to be investigated, as follows:

RQ 4 Can sensed or shared context be used to discover the criteria for contextual recommendation?

In this chapter we have shown that while there can be an overlap between sensed and shared contexts, people do prefer to share context information knowingly, and in ways that do not seem to threaten their privacy or security. This act of sharing can either be integrated, or standalone. It appears from our data that users do not like to share directly sensed context if it is accompanied by a warning, however prepared they are. Further work should investigate the possibilities for abstract contexts which were well-adopted in our experiment.

Current research by Anand and Mobasher (2007) supports a view of interactional context that would change during a recommendation session to ensure a mutual understanding of context between system and user. We have shown here how observation and discussion with a user, in interfaces such as the conversational recommender, can be used to discover the criteria best-suited for them for contextual recommendation.

Chapter 8

Conclusions

In this thesis we have examined how conversational recommenders can be improved and adapted using the wealth of new data becoming available through the Internet. We specifically investigated how traditionally metadata-sparse environments can benefit from conversational techniques, as well as how new contextual information may be interpreted and used to better recommendation. The work we have done examines conversational recommender approaches and extrinsic data; the social trails afforded by relationships and contextual cues that directly affect users. We stated the following primary hypothesis for our thesis in Chapter 1:

Primary Hypothesis Conversational recommenders show great potential to be useful in offering in-situ suggestions and information seeking, but can be made more powerful by harnessing a user's social context.

This chapter marks the conclusion of the thesis. We begin by answering the four research questions we outlined in Chapter 1. We then offer recommendations based on those answers in Section 8.2 before summarizing our contributions to the field in Section 8.3. Finally in Section 8.4 we draw on the work done here to discuss possible future directions.

8.1 Answers to Research Questions

RQ1 How can we create conversational recommenders without intrinsic item knowledge?

In investigating ways to design conversational approaches to recommendation without the traditional overhead of needing item knowledge and needing the user to understand the domain we looked at several things. We evaluated two approaches, one of which was based on collaborative filtering and the other on case-based reasoning, both of which showed an ability to be used by users without domain knowledge, tackling an issue traditionally faced by conversational recommenders. Further to this we found that without resorting to metadata for information filtering the first system could find good items for users faster than traditional interfaces, showing that it was not just easy to converse with but proved effective at finding recommendations, and the second was able to create new items that could be recommended in future through the interface.

We found that by designing systems around capturing initial emotional or reasoned responses to items, rather than experience of the merits of their metadata, we can create a conversation that does not rely on either the user or the system having intrinsic item knowledge. This validated conversational recommenders as able to generalise across and be adapted for modern recommender algorithms.

RQ2 Do conversational recommenders help fulfil a browsing information need?

Answering this question involved querying users about their use of the conversational recommender approach built as part of an exploration into our first research question, as well as an initial exploratory foray into EEG analysis of people using recommendation. We studied user responses to conversational recommendation and found they had no problem stating preferences and traversing a collection to find good items for them. We found that conversational recommenders allow users to browse collections well, even though there are detectable instances where users will reject recommendations before evaluating them. This confirmed that conversational recommenders can offer a successful method of information seeking.

RQ3 Can social relationships inferred from contextual cues prove useful in improving recommendation accuracy?

We looked at the social events surrounding a person's rating to see if there were any detectable clues that preceded their rating which would help predict it more accurately. We found there were many co-occurring factors, with a number that looked promising as data sources for recommender systems. We developed five algorithms to test various strategies for integrating these social signals into recommendation systems and found only the relationships of close friends provided cues that improved recommendation accuracy, with all other relationship tests proving inconclusive. Examinations of authorial influence in the dataset, exploring both authors that were read by and similar to users as well as split by their convergent or divergent interests, were inconclusive. This showed that there are forms of social context that can improve the algorithms behind conversational recommendation.

RQ4 Can sensed or shared context be used to discover the unique criteria for any person's contextual recommendation?

Finally in addressing this question we looked at how users shared context in order to find the forms of context that were acceptable to use. We found users disliked specific sensed context like GPS if accompanied by warning messages at point of collection, but were accepting of completely a short survey to categorise their context in a conversational system. This allowed us to choose the features of social networking profiles on the site Twitter to consider as context that would be evaluated by users. We calculated the F-score (ability to discriminate between users) of each feature, and then trained machine learning algorithms for each of a large number of users and compared what they found important when choosing who to follow. We found that no feature was common enough to be a good context to design for, with each user's own needs representing smaller subsets of the totally available contextual features. This showed that using these features we can discover the unique set of contextual features a person will benefit from. Conversational recommendation, as we have shown with RQ1 and RQ2, can be used in situations where people do not have direct knowledge of how a feature such as context affects them, making it an ideal approach for such a source of information.

Having answered our research questions we found that conversational recommenders are powerful systems to help users browse collections and find good items, and there are both design and algorithmic improvements to conversational recommenders offered by both social and contextual sources. Therefore we are lead to conclude that our primary hypothesis has been confirmed, conversational recommenders, with or without intrinsic item knowledge, can be made more powerful by harnessing a user's social trail and contextual information.

8.2 Recommendations Based on Work

Based on the answers to our research questions we here make some recommendations with regards to conversational recommenders and the sources of data they can use. Note that our findings are specific to recommendation using a conversational interface or making use of social or contextual data; we cannot assume our recommendations will be suitable in other tasks such as list recommendation or personalised search.

Conversational recommendation, as we have shown, is now in a position to be used with collections of items that are not directly comparable using metadata. We recommend conversational approaches to recommendation be considered for more diverse tasks, such as Amazon's entire catalogue or similar collections. Further we recommend that when conversational systems are used no assumption of knowledge on the part of the user is made, rather systems should have the ability to capture gut reactions as we examined. Also it would be prudent, when designing a conversational system such as our collaborative filtering one, to examine the item collection to see how diverse the items are when graphed by average rating and number of ratings. This will allow researchers to decide on an optimal weight to give each answer for partitioning the set. Conversational recommenders support people browsing through a collection, doubly important as we have found as-yet unaccounted for situations where users will reject items no matter what they are shown. This means in order to provide a satisfactory experience users must be given the opportunity, as with conversation, to provide feedback that does not end the recommendation process.

Having looked at the effects of relationships on recommender accuracy we recommend using user-based collaborative filtering algorithms in systems that wish to take advantage of social trails. We showed that though there are a huge number of co-occurring signals and trends not all are easily usable to improve recommendation, so no assumptions can be made when new sources of data become available. Further we recommend that any contextual features can and should be easily tailored to each individual to mirror how they actually use features in their decision making.

8.3 Summary of Contributions

Below we list the main contributions of the scientific investigations we performed in this thesis.

1. We examined a method of eliciting user feedback on items that is compatible with item-based collaborative filtering, allowing conversational recommendation to occur using one of the most common algorithms currently used, around a diverse and dissimilar item collection and requiring no domain knowledge on the part of the user. This expands the utility of conversational recommenda-

tion into all forms of recommendation algorithm currently in use.

2. We showed that conversation can occur between system and user when the system has no intricate knowledge of the domain. This provides a new perspective on the utility of conversation in recommender systems and validates conversation as a method for finding item suggestions in systems even where items are not well described by comparable metadata fields.

3. We showed that conversation can occur between system and user when the *user* has no intricate knowledge of the domain. We designed the process of interacting with the system in a way that offered choices based on popularity, not on giving feedback on specific metadata. Users found this to be easy to respond to based on their gut reactions, regardless of their level of knowledge about the domain, indicating it is possible to create conversational recommenders without a requirement of knowledge, a previously unknown approach.

4. We examined the problem of whether the conversational approach is useful for information seeking. Through user survey and actual EEG signal analysis we showed that conversation is a useful way to browse recommendations except when signals in the brain may indicate a rejection before the fact.

5. We found that close friends can be useful predictors of a user's ratings based on their social trail. While social recommendation has a proven utility we demonstrated that specific social information (the people who someone knows who traditionally have felt similarly about items and shared their feelings *before* that person) can be successfully integrated into a recommender.

6. We showed that experts do not seem to exert influence in the same way friends do. In our examination of influence we looked at the effect of one person's reviews being shared prior to their friends' reviewing of the same item to see if there were detectable trends. We found detectable signals in a number of

categories, including notable trends from close friends and distantly associated people that may be called influence. While one of these signals benefitted recommendation it is interesting to note that experts had no notable cases of influence in the collection.

7. We performed our experiments on publicly available datasets, using three different services covering four domains; movies, running routes, books and microblogging. This ensured our findings were more generalizable. Our datasets, specifically the Goodreads[31] dataset used in Chapter 6 and the Twitter[32] dataset used in Chapter 7 are available on our website[33].

8.4 Future Directions

Having looked at our research questions in depth in this work we have focused on a specific area which, now studied, offers many potential avenues for further exploration. We have identified four areas that show significant potential for scientific discovery in the future and outline them here.

Recommendation as Conversation We have already discussed how recommender systems benefit from new sources of feedback; in this thesis we explored new social feedback and the fact that users value different sources to different degrees. These new sources of information can be used to better understand users, but with conversational approaches we explore tapping the user's knowledge of the situation. The idea that recommendation accuracy is not the only factor in user satisfaction has been discussed since Herlocker *et al.* (2004), but recommenders have yet to take advantage of, as happens in conversation, contextual *feedback*. If a person rates an item or provides implicit feedback (e.g. number

[31] `http://www.goodreads.com`
[32] `http://www.twitter.com`
[33] `http://computing.dcu.ie/~ehurrell`

of plays of a song) that was completely unexpected in a conversational recommender there is an opportunity to ask "*why?*". This opens up an entirely new area to study, how best to recommend armed with the knowledge that a user only watches romantic comedies with their spouse or likes to listen to a certain playlist only on repeat and only in the gym. While sensed context is one method of inferring some of this data we have shown that a conversational approach can procure it directly and unambiguously. Further there can be discussion on a wide range of items leading to interesting research questions around optimal questions for different kinds of items, and what people like to talk about most. This also opens the way for an extended comparison between our approaches, performed on public data, against all other existing and future contributions.

Designing to Support the User In our experiments we focused on the specific task of eliciting and using new information to better recommend items to people. In this we discovered much, including that there are cases when users *do not want to be recommended things.* While this is beyond the scope of our work the implications for study of suitability of recommendation time, and the impact on design of recommenders, warrants further study. Further it is interesting to postulate the best design practices for a digital conversation with users where the aim is to get as much information as possible in order to help the user find good items. This is effectively a new information seeking task born of the ability to exert influence coherently on the recommendation task.

Social Influences in Recommendation We looked at social influences in our work, showing that in a specific understandable way a type of social interaction causes an effect on users and can be accounted for to improve recommendation accuracy. Still to be examined are questions of possible roles of influence and the methods of those roles. For example we could not detect experts as being socially influential, which may seem surprising or it may indicate a vastly different form of influence that we did not detect. Further work is needed to contextualise these social relationships in the same way they are understood in sociological research (such as by Bourdieu (1984), who postulated these relationships were based on expressing similarity or distancing based solely on expression of taste). This could lead to questions of more complex algorithmic accounting for user behaviour and roles in groups, as well as the perceived role they play contrasted against the actual role, which could cause discrepancies in recommendation accuracy.

Context Comparisons We showed in this work that sensed and surveyed context is evaluated differently by different people, showing the potential for systems to account for user difference in viewpoints regarding context. This opens the way to empirically study that which has previously been designed, how context relates to different recommendation tasks, which contextual sensors have no impact on tasks and how to maximally benefit from a smaller number of sensors, i.e. the best sensors to use for a contextually aware holiday or movie recommender.

Appendix

Below are additional tables from our tests in Chapter 6 around Social Context. These tests were to determine the performance of recommender systems taking into account expert influence to improve performance. We tested the approach against a series of common metrics to fully evaluate it.

Next is the full data used in determining representations of context, as described in Chapter 7. These features were scraped from individual tweets to build up a picture of the user who made them, using their contextual information. They were then used as data in the experiment conducted in Section 7.3.

Table I: Convergent vs Divergent Authors Read (P@5)

Test Percent	**AR**	**ARC**	**ARD**
20%	0.00305	0.00364	0.00341
40%	0.01237	0.01272	0.01239
60%	0.04156	0.04391	0.04137
80%	0.18395	0.17872	0.18525

Table II: Convergent vs Divergent Authors Similar (P@5)

Test Percent	**AS**	**ASC**	**ASD**
20%	0.00339	0.00362	0.00347
40%	0.01263	0.01165	0.01248
60%	0.04157	0.04043	0.04106
80%	0.18436	0.19218	0.18030

Table III: Convergent vs Divergent Authors Read (P@10)

Test Percent	**AR**	**ARC**	**ARD**
20%	0.00337	0.00369	0.00372
40%	0.01295	0.01296	0.01251
60%	0.04277	0.04552	0.04328
80%	0.19627	0.19140	0.19783

Table IV: Convergent vs Divergent Authors Similar (P@10)

Test Percent	**AS**	**ASC**	**ASD**
20%	0.00324	0.00367	0.00360
40%	0.01311	0.01199	0.01270
60%	0.04365	0.04439	0.04478
80%	0.19720	0.20134	0.18821

Table V: Convergent vs Divergent Authors Read (Precision)

Test Percent	**AR**	**ARC**	**ARD**
20%	0.12298	0.12269	0.12399
40%	0.08565	0.08695	0.08685
60%	0.05319	0.05306	0.05253
80%	0.00927	0.00974	0.00940

Table VI: Convergent vs Divergent Authors Similar (Precision)

Test Percent	**AS**	**ASC**	**ASD**
20%	0.12333	0.12354	0.12339
40%	0.08580	0.08725	0.08599
60%	0.05234	0.05274	0.05330
80%	0.00933	0.00887	0.00989

Table VII: Convergent vs Divergent Authors Read (Recall)

Test Percent	**AR**	**ARC**	**ARD**
20%	0.00258	0.00261	0.00261
40%	0.00622	0.00632	0.00630
60%	0.01114	0.01097	0.01045
80%	0.01273	0.01305	0.01246

Table VIII: Convergent vs Divergent Authors Similar (Recall)

Test Percent	**AS**	**ASC**	**ASD**
20%	0.00259	0.00266	0.00265
40%	0.00627	0.00640	0.00639
60%	0.01077	0.01069	0.01097
80%	0.01283	0.01249	0.01330

Table IX: Twitter features selected for context (part 1)

Tweets during 12am-1am
Tweets during 1am-2am
Tweets during 2am-3am
Tweets during 3am-4am
Tweets during 4am-5am
Tweets during 5am-6am
Tweets during 6am-7am
Tweets during 7am-8am
Tweets during 8am-9am
Tweets during 9am-10am
Tweets during 10am-11am
Tweets during 11am-12pm
Tweets during 12pm-4pm
Tweets during 1pm-4pm
Tweets during 2pm-4pm
Tweets during 3pm-4pm
Tweets during 4pm-5pm
Tweets during 5pm-6pm
Tweets during 6pm-7pm
Tweets during 7pm-8pm
Tweets during 8pm-9pm
Tweets during 9pm-10pm
Tweets during 10pm-11pm
Tweets during 11pm-12am
UTC offset where the user is
Number of tweets by user
Number of user’s friends
Number of user’s followers
Number of user’s favourite tweets
Number of lists the user appears on
Does their profile use a background image

Table X: Twitter features selected for context (part 2)

What is their default profile image
Are their tweets geo enabled?
Are they verified as who they say they are?
Does the user see media inline
Does the user have contributors enabled
Is the user's account protected?
defaultprofile attribute
istranslator attribute
The twitter client source used to tweet
Profile sidebar fill colour
Profile text colour
Profile sidebar border colour
Profile background colour
Is the user's profile background tiled?
Location
Timezone
User's language
Name of the place the user is currently at
Twitter's URL for the place
Place country
Place type
Place country code
Place id
Place name
The type of geolocation info the user gives
Length of the user's biography
Number of letters in name
Number of capital letters in name
Are the user's name and screen name equivalent?
Screen name length
Number of capital letters in screen name

Bibliography

A. Abdul-Rahman and S. Hailes. Supporting trust in virtual communities. In *System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on*, pages 9–pp. IEEE, 2000.

G. Adomavicius and A. Tuzhilin. Toward the Next Generation of Recommender Systems: a Survey of the State-Of-The-Art and Possible Extensions. *IEEE Trans. on Knowl. and Data Eng.*, 17(6):734–749, jun 2005. ISSN 1041-4347. doi: 10.1109/TKDE.2005.99. URL `http://dx.doi.org/10.1109/TKDE.2005.99`.

G. Adomavicius and A. Tuzhilin. Context-aware recommender systems. *Recommender Systems Handbook*, pages 217–253, 2011a.

G. Adomavicius and A. Tuzhilin. Context-aware recommender systems. In F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors, *Recommender Systems Handbook*, pages 217–253. Springer US, 2011b. ISBN 978-0-387-85820-3.

T. Allen. Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information Within the R&D organization. *MIT Press Books*, 1, 2003.

N. Alon, B. Awerbuch, Y. Azar, and B. Patt-Shamir. Tell Me Who I Am: an Interactive Recommendation System. *Theory of Computing Systems*, 45:261–279, 2009. ISSN 1432-4350. URL `http://dx.doi.org/10.1007/s00224-008-9100-7`. 10.1007/s00224-008-9100-7.

X. Amatriain, N. Lathia, J. M. Pujol, H. Kwak, and N. Oliver. The wisdom of the few: a collaborative filtering approach based on expert opinions from the web. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '09, pages 532–539, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-483-6. doi: 10.1145/1571941.1572033. URL `http://doi.acm.org/10.1145/1571941.1572033`.

S. Anand and B. Mobasher. Intelligent techniques for web personalization. *Intelligent Techniques for Web Personalization*, pages 1–36, 2005.

S. Anand and B. Mobasher. Contextual recommendation. In B. Berendt, A. Hotho, D. Mladenic, and G. Semeraro, editors, *From Web to Social Web: Discovering and Deploying User and Content Profiles*, volume 4737 of *Lecture Notes in Computer Science*, pages 142–160. Springer Berlin / Heidelberg, 2007. ISBN 978-3-540-74950-9.

S. S. Anand, P. Kearney, and M. Shapcott. Generating semantically enriched user profiles for web personalization. *ACM Trans. Internet Technol.*, 7(4), Oct. 2007. ISSN 1533-5399. doi: 10.1145/1278366.1278371. URL `http://doi.acm.org/10.1145/1278366.1278371`.

S. Asch. Opinions and Social Pressure. *Scientific American*, 193(5):31–35, 1955.

D. Athanasopoulos, A. V. Zarras, V. Issarny, E. Pitoura, and P. Vassiliadis. Cowsami: Interface-aware context gathering in ambient intelligence environments. *Pervasive and Mobile Computing*, 4(3):360 – 389, 2008. ISSN 1574-1192. doi: 10.1016/j.pmcj.2007.12.004. URL `http://www.sciencedirect.com/science/article/pii/S1574119207000740`.

O. Averjanova, F. Ricci, and Q. Nguyen. Map-based interaction with a conversational mobile recommender system. In *Mobile Ubiquitous Computing, Systems,* *Services and Technologies, 2008. Ubicomm '08. the Second International Conference on*, pages 212–218, Oct. 2008. doi: 10.1109/UBICOMM.2008.16.

L. Baltrunas, B. Ludwig, and F. Ricci. Context relevance assessment for recommender systems. In *Proceedings of the 16th international conference on Intelligent user interfaces*, IUI '11, pages 287–290, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0419-1. doi: http://doi.acm.org/10.1145/1943403.1943447. URL http://doi.acm.org/10.1145/1943403.1943447.

M. Bank and J. Franke. Social networks as data source for recommendation systems. In W. Aalst, J. Mylopoulos, N. M. Sadeh, M. J. Shaw, C. Szyperski, F. Buccafurri, and G. Semeraro, editors, *E-Commerce and Web Technologies*, volume 61 of *Lecture Notes in Business Information Processing*, pages 49–60. Springer Berlin Heidelberg, 2010. ISBN 978-3-642-15208-5. URL http://dx.doi.org/10.1007/978-3-642-15208-5_5. 10.1007/978-3-642-15208-5_5.

L. Barkhuus and A. K. Dey. Is context-aware computing taking control away from the user? three levels of interactivity examined. In *UbiComp 2003: Ubiquitous Computing, 5th International Conference*, volume 2864, pages 149–156, 2003. ISBN 3-540-20301-X. doi: 10.1007/b93949. URL http://link.springer.de/link/service/series/0558/bibs/2864/28640149.htm. http://www.odysci.com/article/1010112985191527.

N. Belkin and W. Croft. Information filtering and information retrieval: Two sides of the same coin? *Communications of the ACM*, 35(12):29–38, 1992.

N. J. Belkin. Helping people find what they don't know. *Commun. ACM*, 43(8): 58–61, Aug. 2000. ISSN 0001-0782. doi: 10.1145/345124.345143. URL http://doi.acm.org/10.1145/345124.345143.

A. Bogers. *Recommender Systems for Social Bookmarking.* PhD thesis, Tilburg

University, Tilburg, The Netherlands, 2009. URL `http://ilk.uvt.nl/~toine/phd-thesis/`.

A. Borchers, J. Herlocker, J. Konstan, and J. Reidl. Ganging up on information overload. *Computer*, 31(4):106–108, 1998.

P. Bourdieu. *Distinction: a Social Critique of the Judgement of Taste*. Harvard University Press, 1984.

S. Bourke, K. McCarthy, and B. Smyth. Power to the people: Exploring neighbourhood formations in social recommender system. In B. Mobasher, R. D. Burke, D. Jannach, and G. Adomavicius, editors, *RecSys*, pages 337–340. ACM, 2011. ISBN 978-1-4503-0683-6. URL `http://dblp.uni-trier.de/db/conf/recsys/recsys2011.html#BourkeMS11`.

J. S. Breese, D. Heckerman, and C. M. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In G. F. Cooper and S. Moral, editors, *UAI*, pages 43–52. Morgan Kaufmann, 1998. ISBN 1-55860-555-X.

D. Bridge, M. H. Göker, L. McGinty, and B. Smyth. Case-based recommender systems. *Knowledge Engineering Review*, 20(3):315–320, September 2005.

M. Brunato and R. Battiti. Pilgrim: a location broker and mobility-aware recommendation system. In *in Proceedings of IEEE PerCom2003*, page submitted. IEEE Computer Society, 2002.

M. Brunato and R. Battiti. PILGRIM: A location broker and mobility-aware recommendation system. In *Pervasive Computing and Communications, 2003. (PerCom 2003). Proceedings of the First IEEE International Conference on*, pages 265–272, March 2003. doi: 10.1109/PERCOM.2003.1192749.

E. Brynjolfsson, Y. J. Hu, and M. D. Smith. From niches to riches: The anatomy of the long tail. *Sloan Management Review*, 47(2):67–71, Jul 2006.

R. Burke. Interactive Critiquing for Catalog Navigation in E-Commerce. *Artificial Intelligence Review*, 18(3):245–267, 2002a.

R. Burke. Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, 12(4):331–370, Nov. 2002b. ISSN 0924-1868. doi: 10.1023/A:1021240730564. URL http://dx.doi.org/10.1023/A:1021240730564.

I. Campbell. Interactive Evaluation of the Ostensive Model Using a New Test Collection of Images With Multiple Relevance Assessments. *Information Retrieval*, 2 (1):89–114, 2000.

O. Celma. *Music Recommendation and Discovery in the Long Tail.* PhD thesis, Universitat Pompeu Fabra. Department of Information Technology and Communications, 2007. URL http://www.iua.upf.es/%7Eocelma/PhD/doc/ocelma-thesis.pdf.

C.-C. Chang and C.-J. Lin. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):27:1–27:27, May 2011. ISSN 2157-6904. doi: 10.1145/1961189.1961199. URL http://doi.acm.org/10.1145/1961189.1961199.

L. Chen and P. Pu. Users' Eye Gaze Pattern in Organization-Based Recommender Interfaces. In *Proceedings of the 16th international conference on Intelligent user interfaces*, pages 311–314. ACM, 2011.

L. Connaway, T. Dickey, and M. Radford. if it is too inconvenient, i m not going after it.: convenience as a critical factor in information-seeking behaviors. *Library and Information Science Research*, 33(3):179–190, 2011a. ISSN 0740-8188. doi: 10.1016/j.lisr.2010.12.002. URL http://www.sciencedirect.com/science/article/pii/S0740818811000375.

L. S. Connaway, T. J. Dickey, and M. L. Radford. If It Is Too Inconvenient

I'm Not Going After It: Convenience As a Critical Factor in Information-Seeking Behaviors. *Library & Information Science Research*, 33(3):179–190, July 2011b. ISSN 07408188. doi: 10.1016/j.lisr.2010.12.002. URL http://dx.doi.org/10.1016/j.lisr.2010.12.002.

R. L. de Mántaras, D. McSherry, D. G. Bridge, D. B. Leake, B. Smyth, S. Craw, B. Faltings, M. L. Maher, M. T. Cox, K. D. Forbus, M. T. Keane, A. Aamodt, and I. D. Watson. Retrieval, reuse, revision and retention in case-based reasoning. *Knowledge Eng. Review*, 20(3):215–240, 2005.

J. Derrida. *Of Grammatology / by Jacques Derrida ; Translated by Gayatri Chakravorty Spivak.* Johns Hopkins University Press, Baltimore, 1st American ed. edition, 1976. ISBN 0801818419 0801818796 0801818419 0801818796 0801818796.

A. K. Dey. Understanding and using context. *Personal and Ubiquitous Computing*, pages 4–7, 2001. ISSN 1617-4917. doi: 10.1007/s007790170019. URL http://www.cc.gatech.edu/fce/ctk/pubs/PeTe5-1.pdf. http://www.odysci.com/article/1010112991760792.

R. Dhar, S. Nowlis, and S. Sherman. Trying hard or hardly trying: an analysis of context effects in choice. *Journal of Consumer Psychology*, 9(4):189–200, 2000.

W. O. Douglas and K. Jinmook. Implicit Feedback for Recommender Systems. Technical Report AAAI Technical Report WS-98-08, College of Library and Information Services, University of Maryland, College Park, MD20 742, August 1998. URL http://www.aaai.org/Papers/Workshops/1998/WS-98-08/WS98-08-021.pdf.

P. Dourish. What we talk about when we talk about context. *Personal Ubiquitous Comput.*, 8:19–30, February 2004. ISSN 1617-4909. doi: http://dx.doi.org/10.1007/s00779-003-0253-8. URL http://dx.doi.org/10.1007/s00779-003-0253-8.

S. G. Esparza, M. P. OMahony, and B. Smyth. Mining the real-time web: a novel approach to product recommendation. *Knowledge-Based Systems*, 29(0):3 – 11, 2012. ISSN 0950-7051. doi: 10.1016/j.knosys.2011.07.007. URL `http://www.sciencedirect.com/science/article/pii/S095070511100147X`. ¡ce:title¿Artificial Intelligence 2010¡/ce:title¿.

B. Evans, S. Kairam, and P. Pirolli. Do Your Friends Make You Smarter?: an Analysis of Social Strategies in Online Information Seeking. *Information processing & management*, 46(6):679–692, 2010.

D. Fallows and P. I. . A. L. Project. *Search Engine Use.* Pew Internet & American Life Project, 2008.

T. Fawcett. ROC Graphs: Notes and Practical Considerations for Researchers. Technical report, HP Laboratories, 2004. URL `http://www.hpl.hp.com/techreports/2003/HPL-2003-4.pdf`.

A. Felfernig, L. Chen, and M. Mandl. Recsys'11 workshop on human decision making in recommender systems. In *RecSys*, pages 389–390, 2011.

J. Gemmell, T. Schimoler, M. Ramezani, and B. Mobasher. Adapting k-nearest neighbor for tag recommendation in folksonomies. *Intelligent Techniques for Web Personalization & Recommender Systems*, 2009.

M. Göker and C. Thompson. Personalized Conversational Case-Based Recommendation. *Advances in Case-Based Reasoning*, 1898:29–82, 2000.

J. Golbeck and J. Hendler. Filmtrust: Movie recommendations using trust in web-based social networks. In *Proceedings of the IEEE Consumer communications and networking conference*, volume 96. Citeseer, 2006.

J. A. Golbeck. *Computing and Applying Trust in Web-Based Social Networks.* PhD thesis, University of Maryland at College Park, College Park, MD, USA, 2005. AAI3178583.

D. Goldberg, D. Nichols, B. M. Oki, and D. Terry. Using Collaborative Filtering to Weave an Information Tapestry. *Commun. ACM*, 35(12):61–70, dec 1992. ISSN 0001-0782. doi: 10.1145/138859.138867. URL `http://doi.acm.org/10.1145/138859.138867`.

C. Goodwin and A. Duranti. *Rethinking Context : Language As an Interactive Phenomenon / Edited by Alessandro Duranti and Charles Goodwin.* Cambridge University Press, Cambridge [England] ; New York :, 1992. ISBN 052138169 0521422884 521422884. URL `http://www.loc.gov/catdir/toc/cam022/90026850.html`.

Google, Ipsos, and OTX. The smart shopper, sep 2010. URL `http://www.gstatic.com/ads/research/en/2010_TheSmartShopper.pdf`.

G. Groh and C. Ehmig. Recommendations in taste related domains: Collaborative filtering vs. social filtering. In *Conference on Supporting Group Work: Proceedings of the 2007 international ACM conference on Conference on supporting group work*, volume 4, pages 127–136. Citeseer, 2007.

I. Guy and D. Carmel. Social recommender systems. In *Proceedings of the 20th international conference companion on World wide web*, WWW '11, pages 283–284, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0637-9.

K. Z. Haigh, J. R. Shewchuk, and M. M. Veloso. Exploiting Domain Geometry in Analogical Route Planning. *Journal of Experimental and Theoretical Artificial Intelligence*, 9:509–541, 1997. URL `http://www.cs.cmu.edu/~khaigh/papers/khaigh97d.abstact.html`.

U. Hanani, B. Shapira, and P. Shoval. Information Filtering: Overview of Issues, Research and Systems. *User Modeling and User-Adapted Interaction*, 11(3):203–259, 2001.

B. Hanczar, J. Hua, C. Sima, J. Weinstein, M. Bittner, and E. R. Dougherty. Small-sample precision of roc-related estimates. *Bioinformatics*, 26(6):822–830, 2010. doi: 10.1093/bioinformatics/btq037. URL `http://bioinformatics.oxfordjournals.org/content/26/6/822.abstract`.

J. He and W. Chu. A social network-based recommender system (snrs). *Data Mining for Social Network Data*, pages 47–74, 2010a.

J. He and W. Chu. Design considerations for a social network-based recommendation system (snrs). *Community-Built Databases*, page 73, 2011.

J. He and W. W. Chu. A social network-based recommender system (snrs). In N. Memon, J. J. J. Xu, D. L. L. Hicks, and H. Chen, editors, *Data Mining for Social Network Data*, volume 12 of *Annals of Information Systems*, pages 47–74. Springer US, 2010b. ISBN 978-1-4419-6287-4. URL `http://dx.doi.org/10.1007/978-1-4419-6287-4_4`. 10.1007/978-1-4419-6287-4_4.

G. Healy. *An Analysis of EEG Signals Present During Target Search*. PhD thesis, Dublin City University, Department of Computing, 2012. URL `http://doras.dcu.ie/16778/1/FinalThesis.pdf`.

G. Healy and A. F. Smeaton. Optimising the number of channels in eeg-augmented image search. In *Proceedings of the 25th BCS Conference on Human-Computer Interaction*, BCS-HCI '11, pages 157–162, Swinton, UK, UK, 2011. British Computer Society. URL `http://dl.acm.org/citation.cfm?id=2305316.2305346`.

T. Heath. *Information-Seeking on the Web With Trusted Social Networks–From Theory to Systems*. PhD thesis, Citeseer, 2008.

J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl. An Algorithmic Framework for Performing Collaborative Filtering. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '99, pages 230–237, New York, NY, USA, 1999. ACM. ISBN 1-58113-096-1. doi: 10.1145/312624.312682. URL http://doi.acm.org/10.1145/312624.312682.

J. L. Herlocker, J. A. Konstan, and J. Riedl. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, CSCW '00, pages 241–250, New York, NY, USA, 2000. ACM. ISBN 1-58113-222-0. doi: 10.1145/358916.358995. URL http://doi.acm.org/10.1145/358916.358995.

J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, 22(1):5–53, Jan. 2004. ISSN 1046-8188. doi: 10.1145/963770.963772. URL http://doi.acm.org/10.1145/963770.963772.

F. Hu, A. Smeaton, and Y. Sun. An Image Retrieval System Based on Explicit and Implicit Feedback on a Tablet Computer. In *iHCI 2012: Irish Human Computer Interaction Conference 2012, 20-21 June 2012, Galway, Ireland*.

E. Hurrell and A. Smeaton. An Examination of User-Focused Context-Gathering Techniques in Recommendation Interfaces. In *Irish Human Computer Interaction 2012, Galway, Ireland*, 2012.

E. Hurrell and A. F. Smeaton. The benefits of opening recommendation to human interaction. In L. Little and L. M. Coventry, editors, *BCS HCI*, pages 479–484. ACM, 2011.

E. Hurrell, A. F. Smeaton, and B. Smyth. Interactivity and multimedia in case-based recommendation. In G. M. Youngblood and P. M. McCarthy, editors, *FLAIRS Conference*. AAAI Press, 2012. ISBN 978-1-57735-558-8.

P. Ingwersen. Cognitive perspectives of information retrieval interaction: Elements of a cognitive ir theory. *Journal of documentation*, 52(1):3–50, 1996.

P. Ingwersen and K. Järvelin. *The Turn: Integration of Information Seeking and Retrieval in Context (The Information Retrieval Series)*. Springer, 2005. URL `http://www.amazon.com/Turn-Integration-Information-Seeking-Retrieval/dp/904816981X`.

A. Jameson and B. Smyth. Recommendation to Groups. In *The Adaptive Web: Methods and Strategies of Web Personalization*, chapter 20, pages 596–627. 2007. doi: 10.1007/978-3-540-72079-9_20.

K. Järvelin. IR Research: Systems, Interaction, Evaluation and Theories. In *Advances in Information Retrieval - 33rd European Conference on IR Research, ECIR 2011*, pages 1–3, April 2011.

P. Kazienko, K. Musial, and T. Kajdanowicz. Multidimensional social network in the social recommender system. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 41(4):746–759, 2011.

B. P. Knijnenburg and M. C. Willemsen. Understanding the effect of adaptive preference elicitation methods on user satisfaction of a recommender system. In *Proceedings of the Third ACM Conference on Recommender Systems*, RecSys '09, pages 381–384, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-435-5. doi: http://doi.acm.org/10.1145/1639714.1639793. URL `http://doi.acm.org/10.1145/1639714.1639793`.

B. P. Knijnenburg, N. J. Reijmer, and M. C. Willemsen. Each to his own: How different users call for different interaction methods in recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*, RecSys '11, pages 141–148, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0683-6. doi: http://doi.acm.org/10.1145/2043932.2043960. URL http://doi.acm.org/10.1145/2043932.2043960.

Y. Koren. Factorization Meets the Neighborhood: a Multifaceted Collaborative Filtering Model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '08, pages 426–434, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-193-4. doi: 10.1145/1401890.1401944. URL http://doi.acm.org/10.1145/1401890.1401944.

Y. Koren and J. Sill. Ordrec: an ordinal model for predicting personalized item rating distributions. In *Proceedings of the fifth ACM conference on Recommender systems*, RecSys '11, pages 117–124, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0683-6. doi: 10.1145/2043932.2043956. URL http://doi.acm.org/10.1145/2043932.2043956.

Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

C. 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.

S. Lam and J. Riedl. Shilling recommender systems for fun and profit. In *Proceedings of the 13th international conference on World Wide Web*, pages 393–402. ACM, 2004.

H. Lieberman and T. Selker. Out of Context: Computer Systems That Adapt to, and Learn from, Context. *IBM Systems Journal*, 39(3):617–632, 2000. ISSN 0018-8670. doi: 10.1147/sj.393.0617. URL http://dx.doi.org/10.1147/sj.393.0617.

G. Linden, B. Smith, and J. York. Amazon.com Recommendations: Item-To-Item Collaborative Filtering. *Internet Computing, IEEE*, 7(1):76–80, 2003. URL `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1167344`.

F. Liu and H. J. Lee. Use of social network information to enhance collaborative filtering performance. *Expert Systems with Applications*, 37(7):4772 – 4778, 2010. ISSN 0957-4174. doi: DOI:10.1016/j.eswa.2009.12.061. URL `http://www.sciencedirect.com/science/article/pii/S0957417409011075`.

J. M. Lobo, A. Jiménez-Valverde, and R. Real. AUC: A Misleading Measure of the Performance of Predictive Distribution Models. *Global Ecology and Biogeography*, 17(2):145–151, 2008. ISSN 1466-8238. doi: 10.1111/j.1466-8238.2007.00358.x. URL `http://dx.doi.org/10.1111/j.1466-8238.2007.00358.x`.

H. Ma, T. Zhou, M. Lyu, and I. King. Improving recommender systems by incorporating social contextual information. *ACM Transactions on Information Systems (TOIS)*, 29(2):9, 2011.

O. Madani and D. DeCoste. Contextual recommender problems. In *Proceedings of the 1st international workshop on Utility-based data mining*, UBDM '05, pages 86–89, New York, NY, USA, 2005. ACM. ISBN 1-59593-208-9. doi: http://doi.acm.org/10.1145/1089827.1089838. URL `http://doi.acm.org/10.1145/1089827.1089838`.

C. D. Manning, P. Raghavan, and H. Schtze. *Introduction to Information Retrieval*. Cambridge University Press, New York, 2008. ISBN 9780521865715 0521865719. URL `http://www.worldcat.org/search?qt=worldcat_org_all&q=9780521865715`.

G. Marchionini. *Information Seeking in Electronic Environments*. Cambridge University Press, New York, USA, 1995.

G. Marchionini. Toward human-computer information retrieval. *Bulletin of the American Society for Information Science and Technology*, 32(5):20–22, 2006a. ISSN 1550-8366. doi: 10.1002/bult.2006.1720320508. URL http://dx.doi.org/10.1002/bult.2006.1720320508.

G. Marchionini. Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46, 2006b.

P. Massa and P. Avesani. Trust-aware collaborative filtering for recommender systems. *On the Move to Meaningful Internet Systems 2004: CoopIS, DOA, and ODBASE*, pages 492–508, 2004.

P. Massa and B. Bhattacharjee. Using trust in recommender systems: an experimental analysis. *Trust Management*, pages 221–235, 2004.

N. Matthijs and F. Radlinski. Personalizing web search using long term browsing history. In *Proceedings of the fourth ACM international conference on Web search and data mining*, WSDM '11, pages 25–34, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0493-1. doi: 10.1145/1935826.1935840. URL http://doi.acm.org/10.1145/1935826.1935840.

W. Mazzarella. The myth of the multitude, or, who's afraid of the crowd? *Critical Inquiry*, 36(4):697–723, 2010.

K. McCarthy, J. Reilly, L. McGinty, and B. Smyth. On the Dynamic Generation of Compound Critiques in Conversational Recommender Systems. In P. D. Bra and W. Nejdl, editors, *Proceedings of the 3rd International Conference on Adaptive Hypermedia and Web-Based Systems (AH04)*, volume 3137 of *Lecture Notes in Computer Science*, pages 176–184. Springer, 2004a. ISBN 3-540-22895-0. URL http://dblp.uni-trier.de/db/conf/ah/ah2004.html#McCarthyRMS04.

K. McCarthy, J. Reilly, L. McGinty, and B. Smyth. Thinking Positively - Explanatory Feedback for Conversational Recommender Systems. In P. Cunningham and D. McSherry, editors, *European Conference on Case-Based Reasoning (ECCBR-04) Explanation Workshop*, pages 115–124, 2004b.

D. McDonald. Recommending collaboration with social networks: a comparative evaluation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 593–600. ACM, 2003.

D. McDonald and M. Ackerman. Expertise Recommender: a Flexible Recommendation System and Architecture. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 231–240. ACM, 2000a.

D. W. McDonald and M. S. Ackerman. Expertise recommender: a flexible recommendation system and architecture. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, CSCW '00, pages 231–240, New York, NY, USA, 2000b. ACM. ISBN 1-58113-222-0. doi: http://doi.acm.org/10.1145/358916.358994. URL `http://doi.acm.org/10.1145/358916.358994`.

L. McGinty and B. Smyth. Collaborative Case-Based Reasoning: Applications in Personalised Route Planning. In D. W. Aha and I. Watson, editors, *ICCBR*, volume 2080 of *Lecture Notes in Computer Science*, pages 362–376, 2001. ISBN 3-540-42358-3.

L. McGinty and B. Smyth. Shared Experiences in Personalized Route Planning. In S. M. Haller and G. Simmons, editors, *Flairs Conference*, pages 111–115. AAAI Press, 2002a. ISBN 1-57735-141-X.

L. McGinty and B. Smyth. Evaluating preference-based feedback in recommender systems. In M. O'Neill, R. Sutcliffe, C. Ryan, M. Eaton, and N. Griffith, editors, *Artificial Intelligence and Cognitive Science*, volume 2464 of *Lecture Notes in*

Computer Science, pages 209–214. Springer Berlin / Heidelberg, 2002b. ISBN 978-3-540-44184-7.

L. McGinty and B. Smyth. On the Role of Diversity in Conversational Recommender Systems. *Case-Based Reasoning Research and Development*, 2689:1065–1065, 2003.

K. McNally, M. O'Mahony, B. Smyth, M. Coyle, and P. Briggs. Towards a reputation-based model of social web search. In *Proceedings of the 15th international conference on Intelligent user interfaces*, pages 179–188. ACM, 2010.

S. M. McNee, J. Riedl, and J. A. Konstan. Making recommendations better: an analytic model for human-recommender interaction. In *CHI '06 extended abstracts on Human factors in computing systems*, CHI EA '06, pages 1103–1108, New York, NY, USA, 2006. ACM. ISBN 1-59593-298-4. doi: http://doi.acm.org/10.1145/1125451.1125660. URL `http://doi.acm.org/10.1145/1125451.1125660`.

F. Meyer, F. Fessant, F. Clrot, and . Gaussier. Toward a new protocol to evaluate recommender systems. *CoRR*, abs/1209.1983, 2012. URL `http://dblp.uni-trier.de/db/journals/corr/corr1209.html#abs-1209-1983`.

K. Miyo, Y. S. Nittami, Y. Kitagawa, and K. Ohe. Development of Case-Based Medication Alerting and Recommender System: a New Approach to Prevention for Medication Error. *Stud Health Technol Inform*, 129(Pt 2):871–874, 2007. URL `http://www.hubmed.org/display.cgi?uids=17911840`.

B. Mobasher, D. Jannach, W. Geyer, and A. Hotho. 4th acm recsys workshop on recommender systems and the social web. In *RecSys*, pages 345–346, 2012.

M. Montaner, B. López, and J. De La Rosa. A Taxonomy of Recommender Agents on the Internet. *Artificial Intelligence Review*, 19(4):285–330, 2003. URL `http://dblp.uni-trier.de/db/journals/air/air19.html#MontanerLR03`.

M. R. Morris, J. Teevan, and K. Panovich. A Comparison of Information Seeking Using Search Engines and Social Networks. In W. W. Cohen and S. Gosling, editors, *ICWSM*. The AAAI Press, 2010. URL `http://dblp.uni-trier.de/db/conf/icwsm/icwsm2010.html#MorrisTP10`.

A. Muralidharan, Z. Gyongyi, and E. Chi. Social annotations in web search. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, pages 1085–1094, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4. doi: 10.1145/2208516.2208554. URL `http://doi.acm.org/10.1145/2208516.2208554`.

W. Newman and W. M. Newman. Better or just different? on the benefits of designing interactive systems in terms of critical parameters. In *In Designing Interactive Systems (DIS97*, pages 239–245. ACM Press, 1997.

A. Noulas, S. Scellato, N. Lathia, and C. Mascolo. A random walk around the city: New venue recommendation in location-based social networks. In *IEEE International Conference on Social Computing*, 2012.

J. O'Donovan and B. Smyth. Trust in recommender systems. In *Proceedings of the 10th international conference on Intelligent user interfaces*, pages 167–174. ACM, 2005.

J. O'Donovan, B. Smyth, B. Gretarsson, S. Bostandjiev, and T. Höllerer. Peerchooser: Visual Interactive Recommendation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1085–1088, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-011-1. doi: 10.1145/1357054.1357222. URL `http://doi.acm.org/10.1145/1357054.1357222`.

K. Oku, S. Nakajima, J. Miyazaki, and S. Uemura. Context-aware svm for context-dependent information recommendation. In *Mobile Data Management,* *2006. MDM 2006. 7th International Conference on*, page 109, may 2006. doi: 10.1109/MDM.2006.56.

M. O'Mahony, N. Hurley, N. Kushmerick, and G. Silvestre. Collaborative recommendation: a robustness analysis. *ACM Transactions on Internet Technology (TOIT)*, 4(4):344–377, 2004.

U. Panniello, A. Tuzhilin, M. Gorgoglione, C. Palmisano, and A. Pedone. Experimental comparison of pre- vs. post-filtering approaches in context-aware recommender systems. In *Proceedings of the third ACM conference on Recommender systems*, RecSys '09, pages 265–268, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-435-5. doi: 10.1145/1639714.1639764. URL `http://doi.acm.org/10.1145/1639714.1639764`.

M. Park, J. Hong, and S. Cho. Location-based recommendation system using bayesian users preference model in mobile devices. *Ubiquitous Intelligence and Computing*, pages 1130–1139, 2007.

Y.-J. Park and A. Tuzhilin. The long tail of recommender systems and how to leverage it. In *Proceedings of the 2008 ACM conference on Recommender systems*, RecSys '08, pages 11–18, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-093-7. doi: http://doi.acm.org/10.1145/1454008.1454012. URL `http://doi.acm.org/10.1145/1454008.1454012`.

D. Parra and X. Amatriain. Walk the talk: Analyzing the relation between implicit and explicit feedback for preference elicitation. In *Proceedings of the 19th international conference on User modeling, adaption, and personalization*, UMAP'11, pages 255–268, Berlin, Heidelberg, 2011. Springer-Verlag. ISBN 978-3-642-22361-7. URL `http://dl.acm.org/citation.cfm?id=2021855.2021878`.

A. Passos, J. Wainer, and A. Haghighi. What do you know? a topic-model approach to authority identification. *Proc. Of Computational Social Science and Wisdom of Crowds (NIP2010)*, 2010.

M. J. Pazzani and D. Billsus. Content-Based Recommendation Systems. In *The Adaptive Web: Methods and Strategies of Web Personalization*, pages 325–341, 2007. doi: 10.1007/978-3-540-72079-9_10.

M. C. Pham, Y. Cao, R. Klamma, and M. Jarke. A clustering approach for collaborative filtering recommendation using social network analysis. *Journal of Universal Computer Science*, 17(4):583–604, feb 2011. |http://www.jucs.org/jucs$_1 7_4/a_c lustering_a pproach_f or$|.

E. A. Pohlmeyer, J. Wang, D. C. Jangraw, B. Lou, S.-F. Chang, and P. Sajda. Closing the Loop in Cortically-Coupled Computer Vision: a Brain-Computer Interface for Searching Image Databases. *Journal of Neural Engineering*, 8, 2011.

R. Burke. Knowledge-Based Recommender Systems. In *Encyclopedia of Library and Information Systems*, volume 69, page 2000. Marcel Dekker, Inc., Marcel Dekker, 2000.

R. Rafter and B. Smyth. Conversational Collaborative Recommendation — an Experimental Analysis. *Artif. Intell. Rev.*, 24(3-4):301–318, Nov 2005. ISSN 0269-2821. doi: 10.1007/s10462-005-9004-8. URL `http://dx.doi.org/10.1007/s10462-005-9004-8`.

P. Resnick and H. Varian. Recommender Systems. *Communications of the ACM*, 40 (3):56–58, March 1997. ISSN 0001-0782.

P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. GroupLens: an Open Architecture for Collaborative Filtering of Netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, CSCW '94, pages 175–186, New

York, NY, USA, 1994. ACM. ISBN 0-89791-689-1. doi: 10.1145/192844.192905. URL `http://doi.acm.org/10.1145/192844.192905`.

F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors. *Recommender Systems Handbook*. Springer, 2011. ISBN 978-0-387-85819-7. URL `http://dblp.uni-trier.de/db/reference/rsh/rsh2011.html`.

I. Ruthven. Interactive information retrieval. *Annual review of information science and technology*, 42(1):43–91, 2008.

G. Salton and C. Buckley. Improving Retrieval Performance by Relevance Feedback. *Journal of the American Society for Information Science*, 41:288–297, 1990.

B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295. ACM, 2001.

R. Savolainen. Everyday life information seeking. *Encyclopedia of Library and Information Science*, pages 1780–1789, 2010. URL `http://www.informaworld.com/smpp/content~db=all~content=a713532135`.

J. Schafer, J. Konstan, and J. Riedl. Meta-recommendation systems: User-controlled integration of diverse recommendations. In *Proceedings of the eleventh international conference on Information and knowledge management*, pages 43–51. ACM, 2002.

J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen. Collaborative Filtering Recommender Systems. pages 291–324, 2007.

B. Schilit, N. Adams, and R. Want. Context-aware computing applications. In *Mobile Computing Systems and Applications, 1994. WMCSA 1994. First Workshop on*, pages 85–90. IEEE, 1994.

A. Schmidt, M. Beigl, and H. w. Gellersen. There is more to context than location. *Computers and Graphics*, 23:893–901, 1998.

X. Sha, D. Quercia, P. Michiardi, and M. Dell'Amico. Spotting trends: the wisdom of the few. In *RecSys*, pages 51–58, 2012.

U. Shardanand and P. Maes. Social Information Filtering: Algorithms for Automating Word of Mouth. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 210–217. ACM Press/Addison-Wesley Publishing Co., 1995.

E. Shen, H. Lieberman, and F. Lam. What am i gonna wear?: Scenario-oriented recommendation. In *Proceedings of the 12th international conference on Intelligent user interfaces*, IUI '07, pages 365–368, New York, NY, USA, 2007. ACM. ISBN 1-59593-481-2. doi: 10.1145/1216295.1216368. URL `http://doi.acm.org/10.1145/1216295.1216368`.

P. Shenoy and D. S. Tan. Human-Aided Computing: Utilizing Implicit Human Processing to Classify Images. In *Computer Human Interaction*, pages 845–854, 2008. doi: 10.1145/1357054.1357188.

H. Shimazu. ExpertClerk: Navigating Shoppers' Buying Process With the Combination of Asking and Proposing. *Proceedings of the 17th international joint conference on Artificial intelligence - Volume 2*, pages 1443–1448, 2001. URL `http://portal.acm.org/citation.cfm?id=1642194.1642287`.

S. Sinha, K. S. Rashmi, and R. Sinha. Beyond Algorithms: An HCI Perspective on Recommender Systems, 2001.

A. F. Smeaton and C. J. van Rijsbergen. The nearest neighbour problem in information retrieval: An algorithm using upperbounds. In *Proceedings of the Fourth SIGIR Conference*, pages 83–87. ACM, 1981. ISBN 0-89791-052-4.

B. Smyth. Case-based recommendation. In P. Brusilovsky, A. Kobsa, and W. Nejdl, editors, *The Adaptive Web*, volume 4321 of *Lecture Notes in Computer Science*, pages 342–376. Springer Berlin / Heidelberg, 2007. ISBN 978-3-540-72078-2.

B. Smyth. The Sensor Web: Bringing Information to Life . *ERCIM News*, 76:3, 2009.

B. Smyth, L. McGinty, J. Reilly, and K. McCarthy. Compound Critiques for Conversational Recommender Systems. In *Web Intelligence*, pages 145–151. IEEE Computer Society, 2004. ISBN 0-7695-2100-2. URL `http://dblp.uni-trier.de/db/conf/webi/webi2004.html#SmythMRM04`.

E. I. Sparling and S. Sen. Rating: How Difficult Is It? In *Proceedings of the fifth ACM conference on Recommender systems*, RecSys '11, pages 149–156, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0683-6. doi: 10.1145/2043932.2043961. URL `http://doi.acm.org/10.1145/2043932.2043961`.

K. Swearingen and R. Sinha. Beyond algorithms: An hci perspective on recommender systems. In *ACM SIGIR 2001 Workshop on Recommender Systems*, volume Vol. 13, Numbers 5-6, pages 393–408. Citeseer, 2001.

L. Terveen and W. Hill. Beyond recommender systems: Helping people help each other. *HCI in the New Millennium. Addison Wesley*, pages 487–509, 2001.

L. Terveen, W. Hill, B. Amento, D. McDonald, and J. Creter. Phoaks: A system for sharing recommendations. *Communications of the ACM*, 40(3):59–62, 1997.

K. H. L. Tso and L. Schmidt-Thieme. Evaluation of Attribute-Aware Recommender System Algorithms on Data With Varying Characteristics. In *Proceedings of the 10th Pacific-Asia conference on Advances in Knowledge Discovery and Data Mining*, PAKDD'06, pages 831–840, Berlin, Heidelberg, 2006. Springer-Verlag. ISBN 3-540-33206-5, 978-3-540-33206-0. doi: 10.1007/11731139_97. URL `http://dx.doi.org/10.1007/11731139_97`.

K. H. L. Tso-Sutter, L. B. Marinho, and L. Schmidt-Thieme. Tag-Aware Recommender Systems by Fusion of Collaborative Filtering Algorithms. In *Proceedings of the 2008 ACM symposium on Applied computing*, SAC '08, pages 1995–1999, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-753-7. doi: 10.1145/1363686.1364171. URL `http://doi.acm.org/10.1145/1363686.1364171`.

D. Tunkelang. Recommendations As a Conversation With the User. In B. Mobasher, R. D. Burke, D. Jannach, and G. Adomavicius, editors, *RecSys*, pages 11–12. ACM, 2011. ISBN 978-1-4503-0683-6. URL `http://dblp.uni-trier.de/db/conf/recsys/recsys2011.html#Tunkelang11`.

P. Viappiani and C. Boutilier. Recommendation Sets and Choice Queries: There Is No Exploration/Exploitation Tradeoff! In W. Burgard and D. Roth, editors, *AAAI*. AAAI Press, 2011. URL `http://dblp.uni-trier.de/db/conf/aaai/aaai2011.html#ViappianiB11`.

M. Warner. Publics and counterpublics. *Chicago: Duke University Press*, 2005.

P. Wärnestål. User Evaluation of a Conversational Recommender System. In *Workshop on Knowledge and Reasoning in Practical Dialogue Systems IJCAI 2005, 2005*, pages 32–39.

Y. wei Chen and C. jen Lin. Combining svms with various feature selection strategies. In *Taiwan University*. Springer-Verlag, 2005.

M. L. Wilson and m. c. schraefel. Evaluating Collaborative Search Interfaces with Information Seeking Theory. *CoRR*, abs/0908.0703, 2009. URL `http://dblp.uni-trier.de/db/journals/corr/corr0908.html#abs-0908-0703`.

T. Wilson. Models in information behaviour research. *Journal of Documentation*, 55: 249 – 270, 1999a.

T. D. Wilson. Models in information behaviour research. *Journal of Documentation*, 55(3):249–270, 1999b. URL `http://www.emeraldinsight.com/10.1108/EUM0000000007145`.

S. Wold, M. Sjöström, and L. Eriksson. PLS-Regression: a Basic Tool of Chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2):109–130, Oct. 2001. ISSN 01697439. doi: 10.1016/S0169-7439(01)00155-1. URL `http://dx.doi.org/10.1016/S0169-7439(01)00155-1`.

W.-S. Yang, H.-C. Cheng, and J.-B. Dia. A location-aware recommender system for mobile shopping environments. *Expert Systems with Applications*, 34(1):437 – 445, 2008. ISSN 0957-4174. doi: DOI:10.1016/j.eswa.2006.09.033. URL `http://www.sciencedirect.com/science/article/pii/S0957417406002934`.

S. O. Yoon and I. Simonson. Choice Set Configuration As a Determinant of Preference Attribution and Strength. *Journal of Consumer Research*, 35(2):324–336, 2008.

Zhang, Jingjing. Anchoring Effects of Recommender Systems. In *Proceedings of the fifth ACM conference on Recommender systems*, RecSys '11, pages 375–378, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0683-6. doi: 10.1145/2043932.2044010. URL `http://doi.acm.org/10.1145/2043932.2044010`.

C. Ziegler and G. Lausen. Analyzing correlation between trust and user similarity in online communities. *Trust Management*, pages 251–265, 2004.