Chapter 1

Introduction

The birth and proliferation of the Internet has revolutionised the way people interact with and consume information. Resources that were previously difficult and expensive to find are now instantly available; where communication used to be slow, it is now instantaneous and effortless, and tantalising projects such as group-edited encyclopedias are now easily crowdsourced to millions. However, the evolving nature of the web—including the rise of social media—means that the paradigm shift in how people interact online continues to this day. A recent example is the explosive growth of microblogging: the content of web (and, in particular, social network) sites is constantly growing. The single constant aspect of this landscape is that it is ever-changing.

Two themes have emerged from this setting. The first is *cooperation*: many online tasks now take full advantage of interactions between web users in order to help each of them accomplish their own goals. Wikis, blogs, product reviews, and question/answer sites are a handful of examples where users can both contribute (produce) and gain (consume) information: the Internet has blurred the lines between those who create and enjoy content. The second is a battle for users' *attention*. There is now a plethora of sources where content is available; keeping up to date with the latest (movies, news, music, etc) is an established daily challenge.

Recommender systems have emerged in the last decade as powerful tools that people can use to navigate large databases according to their own interests. The engine that underlies these tools is a Collaborative Filtering (CF) algorithm, which is an automated means of ranking content "based on the premise that people looking for information should be able to make use of what others have already found and evaluated" [ME95]. In doing so, these systems capture both of the above themes: users implicitly *cooperate* as the CF algorithm uses each of their ratings and can focus their *attention* on the content that is likely to be of most interest to them. The key insight here is that, while retrieving information from the web tends to be a solitary task, the collective set of experiences can be used to help each individual: recommender systems can evaluate information quality based on the preferences of others with a similar point of view.

In this chapter, we introduce the motivations that seed recommender system research (Section 1.1) and provide a brief overview of the history of recommender system research (Section 1.2); we then define the scope and problem space that we will address in this thesis and outline the contributions that

all subsequent chapters will make (Section 1.3). We close this chapter by listing a set of publications that are related to this thesis.

1.1 Motivating Information Filtering

The most widely discussed motivation for researching recommender systems has remained largely unmodified throughout the 15 years that these systems have been actively studied. It can be described as follows. People's ability to make meaningful use of information (for example, finding and reading interesting news reports) is rapidly falling behind the *rate* at which information is growing (i.e., the number of news reports that are becoming available). The web itself is not only saturated with content, but constantly growing and evolving; people now suffer from the effects of *information overload*. Although this term was first coined by Alvin Toffler in 1970 [Tof70], it continues to describe the difficulty people have when trying to navigate large information repositories—regardless of whether they contain web documents or research articles, e-commerce catalogue products, musicians, movies (and so forth). Recommender systems come to the rescue by suggesting new content to users based on what they have liked in the past [PM97].

Clay Shirky has recently provided an alternative perspective which also substantially motivates research into information filtering [Shi09]. He claims that "we are to information overload as fish are to water," and argues that information abundance has become the norm rather than the problem. Instead, he argues that our focus should be on identifying how we previously filtered information and why those filters, in the face of the information age, are no longer appropriate. For example, publishers—who have the means to produce any book of their choosing—filter their output based on an economic incentive: they are unlikely to sell books of low quality. The advent of the web, however, removed the physical cost of binding a book. In doing so, it eliminated the incentive to publish a narrower range of titles. Furthermore, individuals can now circumvent the publishing houses altogether and directly publish their content online at little to no cost. In other words, we used to filter publications based on (a) money (the economic incentive to sell what is published) and (b) convenience (it was simply too difficult to self-publish). The web has broken the filters we used to rely on by removing these constraints: we now need new ways of filtering published media, and using automated recommender systems to do so may be the tool we are looking for.

The final motivation that we discuss here also takes a different stance to the general context of information filtering. As above, a vast amount of available content is assumed to already exist; the motivation to filter it is that doing so results in heightened user activity (which often translates to increased revenue for web-based businesses). Building tools like recommender systems, that offer personalised views of a web site's content to visiting users, encourages people to actively engage with the site: two thirds of the movies rented by Netflix.com were recommended, Google news recommendations result in 38% more clickthroughs, and 35% of the product sales on Amazon.com were recommended items [CL07]. The motivation to filter information is therefore the fact that tailoring what each user sees to their own needs has been the secret to success of a number of online business.

1.2 Brief History of Recommender Systems

Over the last decade, research into recommender systems has evolved: the particular target scenarios that have been explored have mirrored changes to the way people use the Internet. In the early 1990s, the first filtering system, Tapestry, was developed at the Xerox Palo Alto Research Center [GNOT92]. This system, recognizing that simple mailing lists do not ensure that all users interested in an e-mail's content receive the message, allowed users to annotate e-mail messages so that others could filter them by building queries. This was the first system to capture the power of combining human judgments (expressed as message annotations) with automated filtering, in order to benefit all of the system's users. Similar concepts were later applied to Usenet news by the GroupLens research project, which extended previous work by applying the same principles to the Internet discussion forum, which had become too big for any single user to manage [KMM+97]. The GroupLens project subsequently implemented the MovieLens movie recommender system; the valuable rating data from it was then made available to the wider research community[1], which subsequently shifted focus from news boards toward filtering movies.

The initial success that recommender systems experienced is reflected in the surge of e-commerce businesses that implement them; Schafer *et al.* review and describe a number of mainstream examples [SKR99, SKR01]. The cited sites, like Amazon.com and CDNow.com, implement recommenders to build customer loyalty, increase profits, and boost item-cross selling. More recently, web sites like Last.fm have reaped the benefits of collecting user-music listening habits, in order to provide customized radio stations and music recommendations to their subscribers. The influence, presence, and importance of the recommender system is not only well established, but continues to grow over time.

The widespread commercial applicability of recommender systems is mirrored in the research domain by the extensive breadth of fields that these systems have a presence in. Recommender systems are researched in the context of statistics [AW97], machine learning [CS01], human-computer interaction [PC06, HKR00], social network analysis [MY07], distributed and mobile systems [MKR05, LHC07], agent-based artificial societies [WBS07], computational trust [LSE08, ARH97], and more: it is becoming impossible to capture all of the contributions that are being made to recommender system research. More recently, researchers have explored how content-annotations (tags) can be used to compute recommendations [ZC08], how mobility can be used to recommend social network connections [QC09], disseminate content over mobile networks [QHC07], filter online news [DDGR07] and improve search engine performance [SBCO09].

The most significant recent event related to recommender system research was the announcement of the Netflix prize in late 2006. Netflix—an online DVD rental company from the U.S.—released a dataset of user-movie ratings which, to date, remains the largest publicly available set of user-ratings. They challenged the broader community to outpredict their own system by at least 10% and offered a million dollar reward to the team that was best able to do so. The competition's award itself shows the extent to which web-based businesses value their recommender systems; over 20,000 teams spent 3 years tackling the prediction problem before the winners were announced [Kor09b, TJB09, PC09]. A variety

[1] http://www.grouplens.org/node/73

of lessons were learned throughout the course of the competition; we highlight three here:

- **Matrix Factorisation** emerged early in the competition as a powerful prediction algorithm for CF [Pia07]; it subsequently was consistently used throughout all the leading solutions.
- **Ensemble Methods**. The winning teams did not invest their time in designing accurate predictors; instead, they combined hundreds of individual prediction methods that *together* achieved the target prediction accuracy.
- **Temporal Dynamics**. The Netflix dataset included the date when users input each rating; this data was soon found to be very useful when predicting user tastes, since it reflects the changing bias that users may have [Pot08] or can be incorporated into a wider set of classifiers [Kor09a].

The competition also raised a number of questions, which motivate the work in this thesis.

- **Competition Structure**. Does the structure of the competition (i.e., predicting a hidden set of user ratings) reflect how recommender systems are used in practice? In this thesis, we discuss and propose a novel methodology that more closely reflects the reality of deployed recommender systems.
- **Metrics**. The focus of the competition was accuracy: is this the best way to measure the performance of a recommender system? More importantly, while the leading solutions certainly *predict* ratings well, do they provide better *recommendations*? In this thesis, we examine these points by evaluating collaborative filtering across a number of different dimensions (accuracy, diversity, and robustness).

In the following section, we examine these questions by defining the scope of the research presented in this thesis.

1.3 Problem Statement and Contributions

Recommender systems are built as navigational tools and widely deployed online. In practice, this means that a CF algorithm is implemented and then trained with all the available ratings that the system has for the current content; the algorithm can then be queried to produce recommendations for each user. This process is repeated in a cyclical manner. Why? CF algorithms tend to suffer from very high latency; training an algorithm with the ratings of (potentially millions of) users is a very expensive operation, often requiring exponential space and time, and can thus not be repeated at will (there are, however, a few exceptions [GRGP00]). Recommender systems therefore tend to perform iterative, regular updates (e.g., weekly [Mul06]). Users will not be consistently offered newly computed recommendations, and will have to wait for a system update for their latest ratings to be included in the CF training phase. Since recommendations often elicit further ratings, CF algorithms are iteratively retrained in order for them to have learned from all the data (including any that may have been input since they were last trained).

The traditional research methodology used to design and evaluate CF algorithms, instead, is a two-phase process: researchers measure the performance of an algorithm by first training a given algorithm with a set of ratings and then querying it for recommendations. The problem here is that the research

methodology is static, while deployed systems operate in a cyclical manner: there is a rift between how CF algorithms are studied and how they will be used in practice.

In the following chapters, we address problems that revolve around the central theme of *temporal updates to a recommender system*. The research in this thesis constitutes both *methodological* and *algorithmic* contributions; the former being supported by analysis of large scale rating datasets, and the latter validated with empirical experiments. We decompose problems related to this into three groups: those pertaining to the rating data, evaluating a system that is updated, and securing the robustness of an updating system:

- **Temporal Features of Rating Data**. We report the results of an extensive temporal analysis of three rating datasets (Chapter 3). We draw two main conclusions: (a) CF datasets are subject to a variety of changes over time that are not accounted for when computing recommendations (ranging from dataset growth to customer preference drift) and (b) state of the art similarity measures violate the CF assumption that like mindedness persists between people: they do not produce values that consistently reflect similar people.
- **Evaluating Recommender Systems Over Time**. We define a novel methodology for evaluating the temporal performance of CF algorithms (Chapter 4), based on simulating a number of iterative updates. We accompany this methodology with a number of novel metrics that we use to visualise two facets of CF performance:
  - **Accuracy**. We show how temporal accuracy results provide insight into a facet of performance that would otherwise go unnoticed: accuracy does not improve over time, or with additional ratings. We then propose and evaluate a set of hybrid-switching CF algorithms that keep track of and improve upon their own temporal performance. We show how the same switching strategy can be applied to a mixed set of CF algorithms, or to select and update the k-Nearest Neighbour or Singular Value Decomposition parameters.
  - **Diversity**. We define a novel metric for temporal diversity and show how diversity relates to accuracy. Based on the observations we make when analysing CF temporal diversity, we propose and evaluate a set of algorithms that promote diversity over time (Chapter 5).
- **Securing Recommender System Robustness**. We examine the threats that CF algorithms face from a temporal perspective, and show how attackers who do not factor time into their attack can easily be defeated. We then define how temporal attacks may be conducted and show their effects on a large scale dataset of user ratings. We design and evaluate a series of monitors that can identify when a variety of attacks are taking place. We finally show how attackers may modify their attacks in order to circumvent these defences, and discuss the additional difficulty they will face when trying to do so (Chapter 6).

1.3.1 Timeliness of Research

With the completion of the Netflix prize, collaborative filtering research is reaching an interesting juncture. Striving for accuracy, which has long been the focal point of CF evaluation, has now been pushed

to extreme limits [APO09]. New themes, such as context-aware [ASST05] and mobile [QC09] recommender systems are beginning to emerge, and recommender system methods are now being applied to new domains (such as large-scale software projects [LQF10]). However, the temporal aspect of recommender systems has not been addressed, and *all* of the new application domains assume systems that will be deployed over time. Furthermore:

- While recommender systems are widely used online, there is no insight into how these systems perform as they are updated and users continue rating content.
- The importance of *time* has emerged from the Netflix prize. However, work to date only considers the importance of time in terms of drifting customer preferences [Kor09a]. In this thesis, we consider an alternative (though not mutually exclusive) perspective: how the system performs over time.

We believe that the work in this thesis is timely since it proposes novel methods and metrics to evaluate CF algorithms' temporal performance and provides insights based on empirical evidence that cannot be investigated with current research methods.

1.4 Publications Related To This Thesis

The following publications (and submissions) are related to this thesis:

1. [Lat08a] N. Lathia. Computing Recommendations With Collaborative Filtering. Chapter 2 in *Collaborative and Social Information Retrieval and Access: Techniques for Improved User Modeling*. pp. 23–41. September 2008. IGI Global.
2. [LHC08c] N. Lathia, S. Hailes, L. Capra. Trust-Based Collaborative Filtering. In *Joint iTrust and PST Conferences on Privacy, Trust Management and Security (IFIPTM)*. pp 119–134. July 2008. Trondheim, Norway.
3. N. Lathia, S. Hailes, L. Capra. The Role of Trust in Collaborative Filtering. Under Submission.
4. [LHC08b] N. Lathia, S. Hailes, L. Capra. The Effect of Correlation Coefficients on Communities of Recommenders. In *ACM SAC TRECK*. pp. 2000–2005. March 2008. Fortaleza, Brazil.
5. [LHC08a] N. Lathia, S. Hailes, L. Capra. kNN CF: A Temporal Social Network. In *ACM Recommender Systems (RecSys)*. pp. 227–234. October 2008. Lausanne, Switzerland.
6. [LHC09b] N. Lathia, S. Hailes, L. Capra. Temporal Collaborative Filtering With Adaptive Neighbourhoods. In *ACM SIGIR*. pp. 796–797. July 2009. Boston, Massachusetts, USA.
7. [LHC09a] N. Lathia, S. Hailes, and L. Capra. Evaluating Collaborative Filtering Over Time. In *ACM SIGIR Workshop on the Future of IR Evaluation*. pp. 41–42. July 2009. Boston, Massachusetts, USA.
8. [LHC10b] N. Lathia, S. Hailes and L. Capra. Temporal Diversity in Recommender Systems. In *ACM SIGIR*. July 2010. Geneva, Switzerland.
9. [LHC10a] N. Lathia, S. Hailes, L. Capra. Temporal Defenses for Robust Recommendations. ECML/PKDD Workshop on Privacy and Security Issues in Data Mining and Machine Learning September 2010. Barcelona, Spain.

These appear in this thesis as follows: [Lat08a, LHC08c] and [3] review state of the art approaches to CF (Chapter 2), [LHC08b, LHC08a] investigate the temporal qualities of CF algorithms (Chapter

3), [LHC09b, LHC09a] propose novel metrics and algorithms for predicting ratings over time (Chapter 4), [LHC10b] analyses the diversity of recommendations and evaluates methods to augment temporal diversity (Chapter 5), and [LHC10a] addresses the problem of system robustness (Chapter 6).

There are also a number of other publications that were completed during this time period; while relevant to the broader topics of trust and collaborative filtering, they are not directly within the scope of this thesis:

1. [LHC07] N. Lathia, S. Hailes, L. Capra. Private Distributed Collaborative Filtering Using Estimated Concordance Measures. ACM RecSys 2007. Minneapolis, USA.
2. [Lat08b] N. Lathia. Learning to Trust on the Move. In Joint TIME-SPACE Workshops (IFIPTM). June 2008. Trondheim, Norway.
3. [ALP+09] X. Amatriain, N. Lathia, J.M. Pujol, H. Kwak, N. Oliver. The Wisdom of the Few: A Collaborative Filtering Approach Based on Expert Opinions From the Web. In ACM SIGIR. July 2009. Boston, Massachusetts, USA.
4. [LAP09] N. Lathia, X. Amatriain, J.M. Pujol. Collaborative Filtering With Adaptive Information Sources. In *IJCAI Workshop on Intelligent Techniques for Web Personalization and Recommender Systems*. July 2009. Pasadena, California, USA.

1.5 Summary

In this chapter, we have introduced the generic scenario that recommender systems are best suited to: settings where the *volume* of available information is so great that it exceeds the *ability* that users have to find what they are looking for. Recommender systems are useful since they push content to users without requiring them to formulate explicit queries. Instead, they use the implicit *cooperation* between users in order to rank content for each one of them. We discussed the various motivations that may lie behind building such systems: (a) coping with information overload, (b) replacing filters that no longer work, and (c) making money by encouraging users to interact with their recommendations. We then briefly reviewed the 15 years that recommender systems have been researched and the variety of fields that contribute to it—ranging from statistics to human computer interaction. We placed a particular emphasis on the Netflix prize, since questions that arise from the competition's *structure* and *metric* of choice motivated the work in the following chapters.

In this thesis, we focus on the temporal performance of CF algorithms: we aim to analyse and measure how CF operates over time. In doing so, we endeavour to make both methodological and algorithmic contributions, ranging from a variety of temporal analyses (ratings, similarity, prediction and diversity performance) to a wide range of algorithms to address the accuracy, diversity, and robustness of these algorithms over time. We begin in the following chapter by reviewing state of the art collaborative filtering algorithms.