TRUST NETWORKS FOR RECOMMENDER SYSTEMS

P. VICTOR
C. CORNELIS
M. DE COCK

ATLANTIS COMPUTATIONAL INTELLIGENCE SYSTEMS
SERIES EDITOR | D. RUAN

ATLANTIS PRESS

ATLANTIS COMPUTATIONAL INTELLIGENCE SYSTEMS
VOLUME 4

SERIES EDITOR: **DA RUAN**

Atlantis Computational Intelligence Systems

Series Editor:

Da Ruan, Belgian Nuclear Research Centre (SCK•CEN)
Mol & Ghent University, Gent, Belgium

(ISSN: 1875-7650)

Aims and scope of the series

The series 'Atlantis Computational Intelligence Systems' aims at covering state-of-the-art research and development in all fields where computational intelligence is investigated and applied. The series seeks to publish monographs and edited volumes on foundations and new developments in the field of computational intelligence, including fundamental and applied research as well as work describing new, emerging technologies originating from computational intelligence research. Applied CI research may range from CI applications in the industry to research projects in the life sciences, including research in biology, physics, chemistry and the neurosciences.

All books in this series are co-published with Springer.

For more information on this series and our other book series, please visit our website at:

www.atlantis-press.com/publications/books

AMSTERDAM – PARIS

Trust Networks for Recommender Systems

Patricia Victor, Chris Cornelis, Martine De Cock

Ghent University,
Dept. of Applied Mathematics and Computer Science (WE02),
Krijgslaan 281 (S9),
9000 Gent, Belgium

AMSTERDAM – PARIS

Atlantis Press
8, square des Bouleaux
75019 Paris, France

For information on all Atlantis Press publications, visit our website at:

www.atlantis-press.com

Atlantis Computational Intelligence Systems

Volume 1: Linguistic Values Based Intelligent Information Processing: Theory, Methods, and Applications – Zheng Pei, Da Ruan, Jun Liu, Yang Xu
Volume 2: Computational Intelligence in Complex Decision Systems – Da Ruan (Editor)
Volume 3: Intelligence for Nonlinear Dynamics and Synchronisation – K. Kyamakya, A. Bouchachia, J.C. Chedjou

ISBNs
Print: 978-94-91216-07-7
E-Book: 978-94-91216-08-4
ISSN: 1875-7650

Foreword

Recommendation Systems are gaining tremendous prominence in the digital society, and are fast becoming the bastions of electronic commerce. “Will I like this book?”, “Is this a movie I can see with my kids?’, ‘Which hotel will suit me the best?”: we increasingly rely on the social aspects of the world wide web (WWW) to help us navigate through such questions in our everyday life. We are quick to judge, and even quicker to just imitate our friends and do what they did. The magic potion that casts a spell on us to imitate and even at times make irrational decisions is trust. Trust enhanced recommender systems are designed to help us to form an opinion on matters that are not entirely known to us, or even not known at all.

The social web allows people to express their views to other users of the system. We call the resulting network a social network. There are social networks in which the users can explicitly express their opinion as trust and distrust statements. We refer to these kinds of social networks as trust networks.

The particular focus of this book, infusion of the theory of how online trust networks can be modeled and the utility of these models to enhance the quality of recommendations generated in the online recommendation systems arena is not only groundbreaking and innovative; it is likely to be the central pivot for the next generation of research in social network analysis. Think of any system where humans need subjective judgments from their peers and seers. As you start to read the book, it will be quickly evident that issues explored in this book are the backbone of any such system. Some of these broad issues are: who you know, who you don’t know, who you trust, why you trust them, how does this trust translate, aggregate, and propagate; and how to efficiently and correctly identify key trust figures to enhance the quality of generated recommendations.

As the book notes in its introduction, the year was early 2006, the stage was set with the majority of the developed world focusing on the tremendous excitement that online

communication and collaboration tools on the internet had started to foster. At the same time, the developing world was catching up very fast with the number of people getting access to communication technologies through cell-phones and low-cost internet access points. Electronic commerce was outpacing the brick-and-mortar business in every retail sector and sites such as Epinions.com and Amazon.com were becoming the e-bazaars of the post 9/11 tumulus world where it was safer and easier to shop from home than to face the uncertainty of the physical world. The internet was enabling those that never had a voice to express their opinions freely and without the need for filters enforced by traditional media. YouTube and MySpace were redefining individual reputations and authority in real-life more than someone's social standing in the physical world. The key question on everyone's mind then was: can this online world of inter-human communication be trusted? Can trust be used to infuse confidence in getting the most accurate, relevant and unbiased information about commodities (amazon, ebay), news (twitter), relationships (facebook), etc. Being there is only half the work. Keeping your eyes open to address these needs is the other half. The authors of this book were not only there, they also had their eyes open to see the need for addressing these issues about online trust computation to facilitate online commerce.

Considering the extraordinary ability of this book to make you think, reflect, and to likely influence future research directions in enhancing our understanding of how online social networks behave, I wish you, the reader, a most wondrous journey through the folds of Trust Networks for Recommender Systems. Welcome aboard!

Prof. Ankur M. Teredesai, Ph.D.
Institute of Technology
University of Washington, Tacoma,
USA

Web 2.0 is a massive social experiment, and like any experiment worth trying, it could fail. There's no road map for how an organism that's not a bacterium lives and works together on this planet in numbers of excess of 6 billion. But 2006 gave us some ideas. This is an opportunity to build a new kind of international understanding, not politician to politician, great man to great man, but citizen to citizen, person to person.

Cover story TIME Magazine Dec 25th, 2006. Lev Grossman

Preface

'Person of the Year' is an annual issue of the American news magazine Time, featuring someone or something that "has done the most to influence the events of the year" [135]. The list of the last decade contains famous names such as Barack Obama, Vladimir Putin, Bono, and Bill Gates, just to name a few. However, in 2006, the honor went to 'You' [46]. Yes, indeed. You. If you did not invent a life-saving vaccine, or won the lottery that year and let everyone share in your gains, you are probably wondering what it is exactly that you did to deserve this...

The answer to this question lies with the 'Web 2.0', the umbrella term used most often to refer to the current generation of social web applications. The driving forces behind these applications are collaboration, interaction and information sharing; the key factor to their success being the users themselves, in other words: you and me. Without our enthusiasm and curiosity, the social networking site Facebook[1] would not be so popular; without our movie ratings, the e-commerce giant Amazon.com[1] or the movie rental system Netflix[1] would not be able to recommend us a movie that we are very likely to enjoy; without our hotel reviews, only a few would think of consulting the travel guide TripAdvisor[1] for their next holiday, and so on. The following quotation about the Web 2.0 sums it up nicely:

> It's a story about community and collaboration on a scale never seen before. It's about the cosmic compendium of knowledge Wikipedia and the million-channel people's network YouTube and the online metropolis MySpace. It's about the many wresting power from the few and helping one another for nothing and how that will not only change the world, but also change the way the world changes.

[1] See `www.facebook.com`, `www.amazon.com`, `www.netflix.com`, `www.tripadvisor.com`

The source of the quote is the cover article for the Time issue of December 2006. The authors decided that we had a great influence on that year's events. You and me. Because we "control the Information Age": thanks to our input, millions of users can freely look up information on the online encyclopedia Wikipedia[2], MySpace[2] can make people wonder about other lives, YouTube[2] becomes a way for common people to publish and distribute content online, etc.

But these are only a few examples. In fact, while looking up the Time article in preparation for this book, we came across several of the Web 2.0's success stories: typing '2006 time magazine person of the year' yielded over 18 million results on Google; the fourth and fifth hit were two influential blogs (online journals that are frequently updated) and the second one was a Wikipedia page, the textbook example of a wiki (a website where users can easily add or change content). The first hit was the magazine's web page. On that page you could indicate if you wanted to share the article with your friends (on Facebook), if you wanted to 'retweet' it (via the micro-blogging service Twitter[3]) or 'digg' the article (an application[3] to discover and share content on the web). Clearly, the Web 2.0 experiment has not failed so far. On the contrary, four years after the publication of the Time article, social web applications are alive and very kicking.

Of course, not everything stemming from the Web 2.0 wave is wonderful and useful, and consequently the 2006 nomination caused some controversy. We, too, are skeptical about all these hip and shiny applications/toys/gadgets, but we are also convinced that it has brought us a lot of social applications that we can truly benefit from. In this book, we will focus on one such set of applications, namely social recommender systems. In particular, we will show how trust networks, a specific type of social networks, can enhance the recommendation experience.

This book originated from the doctoral thesis of the first author, which was successfully defended in June 2010. Encouraged by the enthusiastic reports of the committee members, we have decided to publish this book, and make the obtained results available to a larger audience. We are grateful to Etienne Kerre, Paulo Pinheiro da Silva, and Steven Schockaert for their comments and suggestions which have clearly influenced the results in this work. We also would like to thank the external members of the reading committee, Bart D'Hoedt, Enrique Herrera-Viedma and Ankur Teredesai, for their useful suggestions on the first version of the thesis, and Da Ruan for his help with the publication of this book. Thanks also

[2] See `wikipedia.org`, `www.myspace.com`, `www.youtube.com`

[3] See `twitter.com` and `digg.com`

to Epinions.com and CouchSurfing.org for making their social network data available. Finally, we would like to thank the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) for the financial support.

Contents

It's not what you know but who you know that makes the difference.

Anonymous

Chapter 1

Introduction

Although the saying above is an old one, it is surprisingly applicable to the Information Age we are living in now. We are flooded with social networking sites on which we can manage our friends, relatives, or business relations. Some of them are merely used to keep track of our acquaintances, but others can be quite convenient for other purposes too, think e.g. of the business oriented social networking tool LinkedIn or the image hosting website and online community Flickr. Many other useful applications will follow in the next sections.

As will become clear throughout this book, the proverb at the top of this page is especially true for the application that we will focus on. Trust-enhanced recommender systems are designed to help us to form an opinion on matters that are not entirely known to us, or even not at all: 'will I like this book?', 'is this a movie that I can see with my kids?', 'which hotel will suit me the best?', ... Trust-based recommender systems can provide us with personalized answers (or 'recommendations') because they use information that is coming from a social network consisting of people we (may) trust.

1.1 Trust Networks

Social web applications often allow people to express their view on other users of the system. We call the resulting network a *social network*. The relations and/or evaluations between the users come in many flavors: users can add their connections as 'friends' in Facebook, bookmark 'interesting people' in Amazon.com, allow 'followers' in Twitter and 'fans' in Yahoo!Answers[1] (which, as the name suggests, gives users the opportunity to ask and answer questions), etc. Apart from these positive labels, in a large group of users, each with their own intentions, tastes and opinions, it is only natural that also negative evaluation

[1] See `answers.yahoo.com`

concepts are needed. For example, the technology news web site Slashdot[2] lets its users tag each other as 'friends', 'fans', 'foes' or 'freaks', and the political forum Essembly[2] as 'friends', 'allies' or 'nemeses'. A lot of the social web applications rely heavily on the relations between their users, and frequently mine the underlying social network to offer new services. Users of Last.fm[2], e.g., are able to listen to personalized radio stations which are based on the music tastes of their friends. As another example, users of the social bookmarking system Delicious[2] can discover the web pages that their friends like.

In this book, we focus on one type of social networks, namely social networks in which the users explicitly express their opinion as trust and distrust statements. We refer to this kind of social networks as *trust networks*. A popular example is the trust network of the consumer review site Epinions.com[3], a large American e-commerce site where users can write reviews about consumer products and assign a rating to the products and the reviews. The main novelty of Epinions, however, is that users can also evaluate other users, by adding them to their personal web of trust or block list (indicating distrust), based on their quality as a reviewer. Another interesting application is CouchSurfing[3], a large world-wide hospitality exchange network. Users can create a profile and indicate if they are offering sleeping accomodation; other users looking for a couch can then browse through the profiles and try to determine which users are trustworthy enough to be their host (and vice versa). To this aim, CouchSurfing provides several evaluation possibilities, such as leaving references or creating friendship relations. After a couch experience, users can also indicate how much they trust or distrust each other, which constitutes a large trust network among the CouchSurfers.

Forming your own opinion on the users might have been easy when the network was still rather small, but nowadays CouchSurfing contains over one million users, making it increasingly difficult to find the hosts/guests that you would get along with well, let alone the ones that are trustworthy. In the same respect, in Epinions, users may find it overwhelming to form an opinion on a particular reviewer: if there are –very often conflicting– opinions of hundreds of users available, how do you find the users that reflect your tastes the most?

As many trust networks are large, it is very unlikely that all users know each other directly. In other words, the network is unlikely to be fully connected. This means that, if a user a wants to form a trust opinion about an unknown user x, a has to inquire about x with one of its own trust relations, say b, who in turn might consult a trust connection, etc., until

[2] See `slashdot.org`, `www.essembly.com`, `www.last.fm`, `delicious.com`

[3] See `www.epinions.com` and `www.couchsurfing.org`

a user connected to x is reached. The process of predicting the trust score along the thus constructed path from a to x is called trust propagation. Since it often happens that a has not one, but several trust connections that it can consult for an opinion on x, we also require a mechanism for combining several trust scores originating from different sources. This process is called trust aggregation. Propagation and aggregation are the two key building blocks of *trust metrics*, a set of techniques which aim to estimate the trust between two unknown users in the network. Note that the word metric has a different meaning here than the traditional mathematical notion of metric as distance measure in a metric space.

So far, we have not mentioned context and/or goal, although this is an important factor in computing trust estimations and taking decisions based on them: for example, your neighbor might be very good at fixing bicycles, so naturally you would trust him with your flat tire, but that does not imply that you would trust him to baby-sit your six months old daughter. A lot of trust frameworks take into account the trust context, especially when they are general models to be used in several kinds of applications/networks, see e.g. [1, 61]. In this book, we omit the context factor for the sake of simplicity (chapters 2-4) and because we focus on recommendation systems for one type of items only; in other words, we work on recommendations and trust statements that belong to the same domain/context (chapters 6-7).

Omitting the context factor does not harm generality: while an agent in a simplified trust model without context can choose between one or more trust statement types (e.g. trust and distrust in a binary application such as Epinions, or 6 gradual trust levels in CouchSurfing), in a trust application that takes into account context, each one of these possible statements must be accompanied by a context statement. In this respect, we can see a trust connection between two agents as a couple (trust statement,context statement). The operators of chapters 3 and 4 can then easily be applied in context-aware trust applications as well, since propagation and aggregation can only be performed sensibly on trust estimations within the same context.

Obviously, context and goal are also present when defining trust. For example, Gambetta defines trust as a subjective probability that an agent will perform a particular action which affects his own actions [31], and Jøsang *et al.* as the extent to which one is willing to depend on somebody in a given situation [63]. As trust is used in a wide range of application domains, plenty of trust definitions exist. Many of them focus on a different aspect of trust, or stem from a different background (e.g. social sciences versus agent theory): Mui *et al.* see trust as a subjective expectation about an agent's future behavior based on the history of

encounters between the agents [99], while Castelfranchi and Falcone augment Gambetta's definition with a competence dimension [17]. These examples illustrate that there is no consensus on how to define trust. In this book, in which we focus on trust that is explicitly given by the users of a recommender application domain, we adopt the general definition of Jøsang *et al.* We will define trust more precisely when needed to grasp the rationale behind trust operators and their properties.

1.2 Recommender Systems

In the previous pages, we mentioned a lot of Web 2.0 applications: Facebook, MySpace, Twitter, Last.fm, blogs, wikis,... For the remainder of this work, however, we will focus on one specific type of applications, namely *social recommender systems*. Social recommender systems use information about their user's profiles and/or relationships to suggest items that might be of interest to them [121]. Such suggestions (recommendations) can come in many forms: top 10 lists, promotions, 'people who liked movie x also liked movie y', '80% of all people found the following review helpful', etc. And it is certainly not all about movies, cds or books only; also other fields might benefit from a good recommendation system; think for example of research papers, travel packages, courses, and so on.

Good and accurate recommender applications that guide users through the vast amounts of online information are gaining tremendous importance, as the wealth of information makes it increasingly difficult to find exactly what you want or need; all the more because every person has his own preferences. Suppose that you want to go to the movies, but have no idea what to choose: you can surely find a lot of opinions and reviews online, but how do you know which ones are the closest to your tastes and likes? This is where personalized recommendation systems come into play.

From an e-commerce perspective too, the value of a good recommender system cannot be underestimated: Cinematch, the recommender of the American online movie rental system Netflix[4], delivers two third of Netflix's rented movies, Google News[4] recommendations generate 38% click-throughs, and Amazon.com claims that 35% of their sales results from recommendations [72]. Their importance is even more illustrated by the Netflix prize competition, which offered a $ 1 000 000 reward for any recommendation algorithm that is 10%

[4] See `www.netflix.com` and `news.google.com`

more accurate than their own Cinematch[5].

Most widely used recommendation systems are either content-based or collaborative filtering methods. Content-based systems tend to have their recommendation scope limited to the immediate neighborhood of a user's past purchase or rating record. For instance, if you have highly rated a romantic movie with Keanu Reeves, your next recommendation might be a romantic movie or a movie featuring Keanu. The system will continue to recommend related items only, and not explore your other interests. In this sense, recommender systems can be improved significantly by (additionally) using collaborative filtering, which typically identifies users whose tastes are similar to yours (we call them 'neighbors') and recommends items that they have liked. This technique allows for more serendipitous recommendations: you might receive recommendations for movies in a genre that you are not familiar with but that are appreciated by your neighbors, so that there is a good chance that you will like them too.

The advanced recommendation techniques that we will discuss in this book adhere to the collaborative filtering paradigm, in the sense that a recommendation for an item is based on ratings by other users for that item, rather than on an analysis of the item's content. In this sense, as with collaborative filtering systems, they also belong to the class of social recommender systems. More specifically, we will focus on one present-day set of social recommenders, namely *trust-enhanced recommender systems*. The social dimension reaches a whole new level, since trust-enhanced recommenders mine the trust network among their users to offer their services. Such systems incorporate a trust network in which the users are connected by scores indicating how much they trust and/or distrust each other, and use that knowledge to generate recommendations: users can receive recommendations for items rated highly by people in their web of trust (WOT), or even by people who are trusted by these WOT members (through trust propagation and aggregation) etc., yielding more, more accurate, and more personalized recommendations.

1.3 Overview

Trust metrics and recommendation technologies constitute the two pillars of trust-enhanced recommender systems. Trust metrics are covered in detail in chapters 2-4, while chapter 5 deals with the basics of recommender systems. In chapters 6-7 we focus on the intersection of the two fields, viz. trust-enhanced recommender systems.

[5] See `http://www.netflixprize.com`

In chapter 2 we give an overview of existing trust and distrust models and explain their shortcomings. Current models are either not capable of properly handling inconsistency, or cannot differentiate unknown agents from malicious agents. These shortcomings can possibly have a large effect on the (ranking) of trust estimations, recommendations, etc. Therefore, to meet the needs for a framework that can help agents to make better informed (trust) opinions, we propose a new bilattice-based model that preserves valuable provenance information including partial trust, partial distrust, ignorance and inconsistency.

The following two chapters focus on the mechanisms that are needed to predict trust and distrust values in this framework. Chapter 3 covers the propagation problem. Whereas there is a general consensus on how trust can be propagated, the picture gets much more complicated when also distrust is involved. We describe the state of the art of trust propagation, and embark upon the problem of distrust propagation, a research area that has not received much attention so far. We discuss possible distrust propagation strategies, propose and examine a set of propagation operators that exhibit the desired behavior, and illustrate them by investigating propagation patterns in real-world data sets from Epinions and CouchSurfing.

Chapter 4 concentrates on aggregation techniques for trust and distrust values. This field, too, is still in its very infancy. To help in reaching a better understanding of the problem, we propose a set of properties that aggregation operators should fulfill in a (dis)trust context. We demonstrate that the classical aggregation operators for bilattice elements are not always suitable, and therefore propose new families of aggregation operators for trust-enhanced applications. We examine their behavior and show their applicability on data sets from CouchSurfing and Epinions.

The second part of the book deals with the application of trust metrics and their operators in the field of recommender systems. In chapter 5 we cover the recommender basics which are vital for a good understanding of the subsequent chapters. We explain the collaborative filtering mechanism and discuss common evaluation methods and measures (related to coverage and accuracy). We examine the problems of classical recommendation systems — transparency, sparsity, malicious users, cold start users, controversial items — and propose a new detection measure for the latter, which is more suited for evaluation of the corresponding shortcoming.

In chapter 6, we focus on trust- and distrust-enhanced recommendation systems, and show how they can alleviate the problems pointed out in the previous chapter. We provide

a comparative coverage and accuracy analysis of the performance of collaborative filtering and trust-enhanced algorithms for controversial and random items, conducted on data sets from Epinions, and introduce a new algorithm that maximizes the synergy between collaborative filtering and its trust-based variants. Furthermore, to the best of our knowledge, we also provide the first attempt to experimentally evaluate the potential of utilizing distrust in the recommendation process; we investigate the use of distrust as debugger, filter and as an indication to reverse deviations, and its role in the aggregation process for trust-enhanced recommendations.

In chapter 7, we give special attention to the user cold start problem, one of the main difficulties faced by collaborative filtering and trust-enhanced recommender systems. The users of such systems are highly encouraged to connect to other users to expand the trust network, but choosing whom to connect to is often a difficult task. Given the impact this choice has on the delivered recommendations, it is critical to guide newcomers through this early stage connection process. To this aim, we identify several classes of key figures in a trust network, and introduce measures to evaluate the influence of these users on the coverage and accuracy of the recommendation algorithm. Experiments on a dataset from Epinions support the claim that generated recommendations for new users are more beneficial if they connect to an identified key figure compared to a random user; it is indeed who you know that makes the difference.

Seldom, very seldom, does complete truth belong to any human disclosure; seldom can it happen that something is not a little disguised, or a little mistaken.

Emma, 1815. Jane Austen

Chapter 2

Trust Models

Multi-agent systems consist of a large number of intelligent, interactive and (partially) autonomous agents that must cooperate to complete a certain task, often too difficult to solve for an individual agent. Such systems are used in a wide range of applications, ranging from mobile environments [73], over the creation of crowd-related effects for movies[1], to online trading [57]. Multi-agent systems can often benefit from a trust system, especially when the circumstances do not allow for perfect information about the interaction partners' behavior and intentions [117]. They may for example incorporate a trust network to monitor and control the behavior of the agents that participate in a process, think e.g. of an online market place such as eBay. Another nice illustration can be found in [66], in which a trust network is used to alleviate the problem of corrupt sources in peer-to-peer file-sharing networks by keeping track of the peers' trustworthiness. With the advent of the Semantic Web [12], even more applications and systems will need solid trust mechanisms. The Semantic Web is an extension of the current web where content is annotated (see RDF[2] and OWL[3]) such that machines and computers are able to understand its meaning and reason with it. Hence, since more and more intelligent agents will take over human tasks in the future, they also require an automated way of inferring trust in each other, see for instance [123].

Nowadays, effective models already play an important role in many Web 2.0 applications. Question answering systems can compute trust indications along with the answers based on how much trust the user puts into certain sources [153], recommender systems can produce suggestions more tailored to the users' tastes (chap-

[1] Massive Software, see `www.massivesoftware.com`

[2] Resource Description Framework, see `www.w3.org/TR/rdf-primer`

[3] Web Ontology Language, see `www.w3.org/TR/owl-features`