Combining Reputation and Collaborative Filtering Systems

Maria Tkatchenko

University of British Columbia
201-2366 Main Mall
Vancouver, B.C.
V6T 1Z4

Software Practices Lab
1-604-822-1290

tkatch@cs.ubc.ca

ABSTRACT

In this paper, we explore the possibility of using a combination of reputation and collaborative filtering systems to create a more robust and accurate system. The main target of such a system would be an e-commerce website or on-line auction setting.

The main mechanism this proposal presents centers on using reputations to improve recommendations made in a collaborative filtering setting. We also briefly discuss an approach which uses collaborative filtering to improve reputation systems, as a suggestion for future work.

In the case of the main mechanism we predict noticeable improvements in the accuracy of predictions, because of the additional weight provided by the extra information we take into account.

We mention a number of problems with existing reputation and collaborative filtering systems, and discuss how the proposed system would deal with the same kinds of issues.

General Terms

Algorithms, Design, Economics, Human Factors, Theory.

Keywords

Collaborative filtering systems. Reputation systems.

1. INTRODUCTION

In this paper, we mainly explore the possibility of using reputation systems to improve the performance of collaborative filtering (recommendation) systems[1]. In particular, we are interested in discovering which features of one system can be used to cover the current shortcomings of the other, in order to create a more complete system for users to interact with.

As e-commerce has expanded and become more accepted in the last decade, it has become more important for websites to cater to the demands of customers more efficiently and accurately. In particular, most users expect to be able to apply their real-life experience in their online interactions, as well.

For example, in real life you would not buy an item from a person you knew to be untrustworthy and unlikely to deliver the good, but how are you to know the trustworthiness of an anonymous seller on-line? Reputation systems have been introduced to deal with this problem, where a seller's reputation is built up not through word-of-mouth, but through explicit ratings that buyers assign to their transactions with the seller. Future customers have access to the aggregate of these ratings, and can use them to evaluate a seller's reputation.

Similarly, in real life, people in your social circle often make suggestions as to what books you should read or what items you should buy, based on their experiences with those products. Collaborative filtering systems have been designed to accomplish similar recommendations automatically, by finding groups of individuals who have similar tastes to yours and recommending the items they also found useful. Similarity of tastes is determined based on previous interactions.

This work concentrates on merging the two kinds of systems in hopes of improving their performance. For example, individualized reputations can be improved by tying your personal reputation to the reputations of other users, based on the similarity of ratings both of you gave to other agents or products. Collaborative filtering systems can be improved by assigning reputations to users, perhaps based on the quality of their previous recommendations, or the similarity of their ratings to items that you have rated. If the opinions of more reputable agents are given more weight when making recommendations, this will result in more precise suggestions being made by the system.

A system combining the ideas of both recommendation and reputation systems will provide better support for user interactions, as often the interactions in both of the directions described are needed. In terms of a recommendation system improved using reputation, we believe we can improve both the quality of recommendations that are made, and the range of conditions under which recommendations still make sense.

2. BACKGROUND

This section will introduce the reader in more detail to both reputation systems and collaborative filtering systems, and discuss some of the work that has been done in each of these areas. It will then provide a more detailed discussion for the reasoning behind our proposal to combine the two kinds of systems, and list a number of benefits that could be achieved with a combined system.

[1] We will use the terms collaborative filtering and recommendation systems interchangeably, as we have seen done in other literature surveyed.

2.1 Reputation Systems

Reputation, in general, refers to a certain characteristic of an agent, often related to his trustworthiness, that other agents with whom he interacts (his partners) ascribe to him. The reputation of an agent is most often based on ratings other agents give him after their interactions with him, and are usually calculated through some sort of a global aggregation of all the agents' ratings. In addition, the reputation of an agent can be calculated either explicitly from other agents' ratings of him, or implicitly from the correlations between the ratings this agent and all the others give to objects. There are a number of variations in the interpretation of reputation in the context of multi-agent systems, each of which contributes to its general intuitive meaning.

Reputation reporting systems have become increasingly popular in the last few years with the rise of e-commerce, and studies have shown that seller reputation significantly influences prices in on-line auctions [7, 8]. Houser and Wooders [7] define reputation as "propensities to default" – in other words, for a buyer this is the likelihood he will deliver payment if he wins, while for the seller it is the likelihood that he will deliver the item once payment is received.

Raub and Weesie [9] note that reputations can only emerge in continuing games, in particular if information about an actor's behaviour in one of his relations spreads to his future partners. The authors conclude that reputation can play an important role in achieving equilibria where none were possible before. However, their model neglects the cost of information flow between agents, as well as any costs associated with supply of information, which are unrealistic assumptions in the case of most existing e-commerce applications.

Economists have also applied game theoretic approaches to the study of reputation. In repeated games, which are used to simulate user interactions, reputations of players are linked to the existence of cooperative equilibria in the game [8]. The Folk Theorem is one work that deals with such equilibria.

Chen and Singh [3] provide a sound algorithm for the computation of reputations, as well as an analysis of the state of reputation reporting in on-line systems. The distinguishing characteristic of their algorithm is that the agents' reputations are used as weights in the computation of the rating of an object, and in return the set of all the ratings given to objects evaluated by an agent affects that agent's reputation.

More recent work has also dealt with the issue of separating the various aspects of reputation. For example, Zacharia and Maes [12] deal with distinguishing global from personalized reputation, where the reputation seen is dependent on who makes the query. Sabater et. al. [10] divide reputation into individual, social, and ontological, where in the latter two cases the reputation is based on the social group an individual belongs to or the context in which the query is made, respectively.

Tennenholtz [11] looks at reputation systems from the point of view of social choice, and provides a series of results for various flavours of reputation systems. He categorizes reputation systems by the kind of feedback they involve (positive, negative, or both), and shows for each one that it is impossible to achieve a social choice rule that satisfies generality, transitivity and weak monotonicity.

Dellarocas [4] looks at analyzing the efficiency of markets such as eBay. The reputation mechanism is used here to determine whether sellers advertise the quality of their goods truthfully, in environments where buyers inherently receive asymmetric information about seller behaviour. One of his main results is that efficiency can be achieved when buyers are (i) lenient when rating a seller after an interaction, and (ii) strict when evaluating a seller profile before an interaction. More interestingly, Dellarocas shows that better assessments can be made based on the negative ratings than on the positive ones.

In a later work Dellarocas discusses an incentive mechanism for auctions which maximizes average social welfare [5]. Provided buyers obey a simple binary rating scheme, a schedule of fees and rewards for sellers can be derived which induces sellers to produce at their best quality, and truthfully announce intended quality levels to buyers.

2.2 Collaborative filtering systems

Collaborative filtering systems allow one to make personalized recommendations based on computed similarities between an individual's prior preferences and those of other users. Recommender systems, based on collaborative filtering, help users address the challenges of information overload in on-line marketplaces. Collaborative filtering recommenders use opinions of other agents to predict the value of items for each agent in the community [6].

Collaborative filtering systems often depend either explicitly or implicitly on the ratings that agents give to items they have purchased previously. Then, for a given agent A, a recommendation is computed by finding other users whose preferences for items A has rated correlate most closely to those of A, and recommending other items that these users have rated highly. Thus, for example, personalized purchase recommendations on a website can increase the likelihood of the customer purchasing the item, when compared to unpersonalized recommendations [1].

One problem with collaborative filtering systems, noted in Canny's work [2], is that they do not exactly correspond to how recommendations are usually made in social settings. For example, many people would defer to more advanced individuals with more expertise in an area when looking for a recommendation in that area. Thus, they often seek recommendations from "more advanced peers who are *unlike* them" [2]. Canny presents cooking as an example, where, if you're looking for a good recipe, you will want a recommendation from that specific community, and not your own peers or even the population as a whole. This is one of the more serious drawbacks of collaborative filtering systems, which we will attempt to address in our proposal.

In addition to mentioning a number of such disadvantages with traditional collaborative filtering systems, Canny explores the issue of privacy. In particular, the question of how to hide individual user ratings and still obtain good recommendations. He presents a distributed collaborative filtering algorithm which computes recommendations based on encrypted rating data, using a set of trustworthy peers to perform the calculations.

Good et. al. have looked at combining collaborative filtering with personal agents in order to improve recommendations [6]. They say collaborative filtering focuses on identifying users with similar tastes and using their opinions to recommend items, whereas information filtering focuses on analyzing the item content and developing a personal profile for each user. The authors show that using collaborative filtering to create personalized combinations of sets of agents gives better results than either individual agents or other combination mechanisms. The results were obtained by using the user ratings from the MovieLens system: the system was used to predict ratings of a target user after training on a set of ratings taken from the system, and then the authors compared these results with the user's actual ratings.

Burke [1] looks at collaborative filtering systems which use multi-dimensional or semantic ratings, where the system also gets information about the reason behind a preference. He addresses the question of how much improvement can be expected from adding collaborative filtering to an existing knowledge-based recommender system. The results presented show that the heuristic technique proposed in the paper is clearly superior to those that don't use any heuristic information. These results are especially significant in terms of accuracy, where high accuracy was achieved even for small training sets.

2.3 Why combine the two systems

In large part, our work is motivated by Canny's discussion of the disadvantages of collaborative filtering systems [2]. In particular, he points out that, in order for recommendations to achieve a higher level of precision, they require diffusion of ratings from the more advanced peers. In particular, individuals often seek recommendations from more advanced peers who are not necessarily like them. Collaborative filtering systems, on the other hand, only take into account the recommendations based on the ratings of peers who are most similar to you.

Collaborative filtering systems provide reasonable recommendations, but a lot of work still needs to be done on improving their precision. Recommendations are often done by aggregating the users into groups based on their preferences for sets of items, and then making recommendations to each user based on what other agents in the group preferred, but the individual hasn't tried. If groups were determined, in part, by calculating the reputations of users in a given category, then the recommendations for that category could be made more precise due to the virtue of these users being more knowledgeable about items in that category.

Finally, it is interesting that we have not discovered much work directly related to both collaborative filtering and reputation. Both areas look at large groups of users and products, where ratings are taken into account and used to evaluate the suitability of an agent (in the case of reputations) or product (in the case of recommendations). The two areas look like a good fit, and we thought it would be an interesting experiment to see what can emerge from combining the two systems, in addition to learning more about both of them over the course of this project.

3. PROPOSAL

Given the justification in the previous section, here we will describe two mechanisms. The first mechanism, for applying reputation to collaborative filtering systems, we will discuss in more detail. We will just touch on the mechanism for using collaborative filtering to make improvements in reputation systems.

3.1 Improving recommendation systems

Previous work [6, 12] has noted that while recommendation systems perform well on individual agents/items, they are often not very accurate in global environments. As mentioned in Section 2.2, one problem with pure collaborative filtering systems is that they provide recommendations based on the ratings of users most similar to you, while in some cases you may want the recommendations to be based on the ratings of users who are quite different from you, but experts in the field in which you are seeking a recommendation. Thus, a reputation system can be used to distinguish groups of users based on their expertise in a given area, and base the recommendations on these reputations when you are looking for items you have not looked at before.

If a reputation system can be used to improve the quality of the correlations of ratings, then recommendations can be made more precise. In this section, we will first provide a discussion of how a reputation can be calculated for a given user, and then how these reputations can be applied to making recommendations.

3.1.1 Calculating reputations

We can think of most users of on-line auction or e-commerce websites as having a number of given categories of interest. They will often buy goods that belong to these categories, more rarely may look at goods in related categories, and very rarely will look at goods that belong to categories that are completely unrelated. Thus, for example, I may be interested in computer textbooks, pictures, and cameras (photography is my hobby), but never look at furniture or pet supplies because they are irrelevant to me at the moment.

The categories that an agent most often buys from can be seen as his areas of *expertise*, or authority. By evaluating multiple items from a given category, he is likely to develop experience in rating the items in this category. More importantly, if he often buys goods from this category, it can be assumed that he has some knowledge of the products in the category, and thus can quite accurately estimate the quality of the good, since he has a large number of items to compare against.

However, an agent's reputation cannot depend only on the number of ratings he has made in a given category. If that were the case, one can imagine a subversive agent building up a reputation simply by rating many items in a single category, but giving low ratings to competitor products and high ratings to products that he wishes to encourage other people to buy, regardless of the merits of the product. Thus, the reputation of an agent in a given category must also depend on how closely his ratings match the ratings given by other users.

3.1.2 Applying reputations to recommendations

Once reputations have been calculated for each user in each relevant category, we can use these to improve the quality of recommendations that the system makes.

However, we must also be careful in assigning weights to an agent's reputation in each of the categories. Some sort of scaling values need to be devised to account for the weigh each of your categories gets. Most intuitively, for the categories in which you have the most expertise, your rating should receive more weight, compared to other people with less expertise.

Until now we have talked about expertise in qualitative terms. But how can we quantify this key concept?

The expertise of an agent in a given category should be a function of at least two things: the number of ratings he has made in the category, and the correlation of his ratings with ratings of other agents who are also knowledgeable in this category.

With respect to the first factor, the number of ratings in this category should be expressed as a fraction of the total number of ratings the agent has made. This is done to give similar weighting to the ratings of agents A and B, for example, where A has made 100 ratings and B 30, but both have made 40% of their ratings in category X, and so should have a similar level of expertise in this category.

However, it may be the case that the total number of ratings may have to be factored into the above calculation, as well, to account for truly prolific agents who are super-users in certain categories. However, here we run into the question of whether the agent is truly a super-user, or simply someone making many ratings in random categories in order to build up a reputation quickly. We leave this as future work, but provide a number of suggestions for addressing this problem later in the paper.

To counter this threat, we need to take into account how well the ratings an agent gives to various products correlate with ratings the other agents assign to the same item. For example, high positive correlation with the more influential users (the experts) will give strong positive feedback to your rating, while a high negative correlation with non-experts will give a weaker negative feedback. Since previous work has shown that negative ratings are actually more accurate in providing assessments [4], we will apply both positive and negative measurements to figure out the correlation portion of the weight. Possible issues of continued expert dominance are briefly addressed in Section 4.5.

3.2 Improving reputation systems

In this section, we will give a brief sketch of how we think concepts from a collaborative filtering system can be used to improve a reputation system.

The work of Chen and Singh [3] appears to be a very good starting point for broadening the scope of information that is taken into account when calculating individual reputations. They, for example, aggregate an agent's overall reputation from his reputation in various categories, which in turn are calculated based on the similarity of the agent's ratings to ratings by other agents.

However, we can go even further, using collaborative filtering in order to better approximate an agent's true reputation. As mentioned before, collaborative filtering systems compute recommendations by computing similarities between your preferences and those of other people. We can apply this method of computing recommendations to computing reputations, as well. In particular, we can weight the ratings that an agent gives you based on their similarity to other agents' ratings, as well as the rater's own reputation. Of course, the same privacy and dominance concerns apply here as in the case of an improved recommendation system. The discussion in the next section applies to this proposal, as well.

4. Evaluation

Due to the time-frame and scope of this project, we did not implement a system to test out the proposed mechanism, so no concrete quantitative results are possible. However, what we chose to do is to explore the more obvious problems of existing approaches, and describe ways in which the proposed system would address them.

4.1 Calculating reputations

One issue we did not consider when talking about calculating reputations in the previous section is that of global versus personalized reputations.

It is possible to have a single reputation value for each agent that is globally available to anyone who requests it. It would also be possible, however, for each agent to keep a set of personalized reputations for all the other agents in the system. This would depend not only on the globally calculated reputation (which is based on an agent's expertise in a given category), but also on the interactions you have had with that agent. If, for example, the recommendations based on that agent's preferences were not useful to you despite his expertise, you may want to privately downgrade that agent's expertise. However, this would require for the system to reveal (i) the information about agents whose preferences were used in making a recommendation, and maybe (ii) the system's mechanism for calculating the recommendations. While (ii) is probably not a great concern if a mechanism is strategy-proof, (i) would definitely not sit well with many of the users concerned with privacy. This issue is also touched on in Section 4.4.

4.2 Number of ratings

One problem noted in [2] is that collaborative filtering systems break down when there are few ratings for an item of interest. The majority of agents rate few objects, so much of the time no accurate correlations can be found between their preferences. Canny even goes so far as to suggest that you have to drop items with few raters since they are simply not useful in making accurate predictions.

The proposed mechanism, however, can improve the matter somewhat. If there are at least a couple of expert users for a given category, then their preferences can simply be listed as recommendations when an agent wishes to see recommendations for that category. As the number of agents who have rated in a given category increases, the experts' opinions get proportionately less weight, since they are no longer the sole agents with knowledge on the subject.

4.3 Collecting information

One possible issue with this system, as with any other system that relies on information gathered about the agents, is that users might be unwilling to have the system collect such information as usage statistics and ratings. However, over the last few years people seem to have gotten used to the idea that reputation and recommendation systems on different websites collect this kind of information. Moreover, the idea that such systems benefit from this kind of feedback has been demonstrated time and again, as they've improved through new kinds of algorithms being developed based on the information collected.

Both rating and usage statistics are necessary for the mechanism described in Section 3.1, since they allow us to obtain more precise recommendations. We believe that users will accept this potential loss of a small amount of privacy in exchange for much more accurate recommendations, especially considering that much of this information is already collected.

4.4 Privacy concerns

One of the major concerns with such systems, discussed, for example, in [2], is that of privacy. Currently, the proposed mechanism doesn't employ any information-hiding mechanisms. However, it would be possible to adapt Canny's distributed collaborative filtering approach to the proposed recommendation system. In this case, the information passed between the peers would also have to include the reputation information, as well. One issue that will need to be addressed in this case is whether the reputations will be globally accessible, or whether each agent will have to maintain a separate list of reputations in order to calculate the recommendations. The issue of global versus private reputations is discussed in Section 4.1.

4.5 Expert dominance

One possible concern in the case of using reputations in recommendation systems is that the opinions of the more expert users may completely overwhelm the opinions of all the other users, when in reality the latter are more right. One can imagine this in the case of established and reputable agents who are out of touch with recent developments (which is very likely in a fast-moving field such as software engineering, or bioinformatics, for example), and whose ratings thus don't reflect the true nature of products. One way to improve this situation is to periodically survey the set of all the recent ratings in the system. If the opinions of expert users consistently differ from the ratings of a significant portion of other agents, then the assessment of their expertise needs to be downgraded to reflect these changes.

5. Future work

One interesting extension of this approach would be to actually implement the mechanism described in the proposal. In that way, the system can be tested out in an on-line setting with real users, and reveal any pitfalls that may result from agents not assigning rankings to objects or other agents in a way that is completely rational.

Given the time limitation of this project, not all the details of the algorithm have been fleshed out completely. The next step would be to explore the algorithm discussed in more detail, and discover any boundary conditions which may cause it to perform poorly, if any exist. In particular, it would be interesting to develop the mechanism mentioned in Section 3.2 to the same level as the mechanism discussed in Section 3.1. In addition, scaling values that determine the importance of each of the agent's rated categories in calculating the reputation value need to be determined. This can probably best be accomplished by implementing the system, and using the values obtained by running it with a set of real users to find the values that achieve the best recommendations. Another possibility is to do what Good et. al. did in [6], and use an existing set of ratings to train the system, and then compare the predictions of the system with the actual ratings given by the users, converging to the scaling values in that way.

One issue that this paper doesn't address is that of the applicability of the proposed system. For example, it would be interesting to determine whether there are any areas where the mechanism can't apply, or performs worse than either of the original mechanisms.

Section 3.2 glosses over the issue of improving reputation systems using collaborative filtering. A natural extension of this work would be to explore this issue in more detail, building a corresponding system to evaluate the ideas. However, we chose to focus on improving recommendation systems with reputation because (i) we are more interested in collaborative filtering systems, and (ii) recommendation systems currently seem to require more work.

6. Conclusion

In this paper, we propose an approach which merges reputation systems and collaborative filtering. There are two possible applications of this approach. First of all, reputation systems can be used to enhance recommendation systems, by assigning more authority to agents in categories where (i) they have made many accurate ratings, and (ii) their ratings correlate closely with the ratings of other users who satisfy (i). On the other hand, collaborative filtering systems can be used to enhance reputation systems, whereby other agents' ratings of you are weighted both by their reputation and similarity to other agents' ratings.

The next step in exploring this proposal would be to develop and test a system which implements this mechanism. Due to time constraints, this was not included in the scope of this work. However, we discuss a number of limitations of existing recommendation and reputation systems, and suggest ways in which the proposed system would improve on these. In particular, we look at privacy concerns and the issue of the number of ratings required for accurate recommendations, as well as the related question of unwanted dominance of expert opinions, even when they may be out of touch. We also suggest a number of possible future directions to explore, based on a number of things we mention in the proposal but do not go into any detail on.

7. REFERENCES

[1] Burke, R., Semantic ratings and heuristic similarity for collaborative filtering. in *Proceedings of the 17th National Conference on Artificial Intelligence*, (2000).

[2] Canny, J., Collaborative Filtering with Privacy. in *IEEE Conference on Security and Privacy*, (2002).

[3] Chen, M. and Singh, J.P. Computing and using reputations for internet ratings *Proceedings of the 3rd ACM conference on Electronic Commerce*, ACM Press, Tampa, Florida, USA, 2001.

[4] Dellarocas, C., Analyzing the Economic Efficiency of eBay-like Online Reputation Reporting Mechanisms. in *Proceedings of the 3rd ACM Conference on Electronic Commerce*, (2001).

[5] Dellarocas, C., Efficiency through Feedback-contingent Fees and Rewards in Auction Marketplaces with Adverse Selection and Moral Hazard. in *3rd ACM Conference on Electronic Commerce*, (2003).

[6] Good, N., Schafer, J.B., Konstan, J.A., Borchers, A., Sarwar, B., Herlocker, J. and Riedl, J., Combining Collaborative Filtering with Personal Agents for Better Recommendations. in *Proceedings of the National Conference of the American Association of Artificial Intelligence*, (1999).

[7] Houser, D.E. and Wooders, J. Reputation in Auctions: Theory, and Evidence from eBay, http://bpa.arizona.edu/~jwooders/ebay.pdf, 2000.

[8] Mui, L., Mohtashemi, M. and Halberstadt, A., Notions of reputation in multi-agents systems: a review. in *Proceedings of the first international joint conference on Autonomous agents and*

multiagent systems: part 1, (Bologna, Italy, 2002), ACM Press, 280-287.

[9] Raub, W. and Weesie, J. Reputation and Efficiency in Social Interactions: An Example of Network Effects. *The American Journal of Sociology*, *96* (3). 626-654.

[10] Sabater, J. and Sierra, C., REGRET: A reputation Model for Gregarious Societies. in *4th Workshop on Decaption, Fraud and Trust in Agent Societies*, (2001).

[11] Tennenholtz, M. Reputation systems: an axiomatic approach *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, AUAI Press, Banff, Canada, 2004.

[12] Zacharia, G. and Maes, P., Collaborative Reputation Mechanisms in Electronic Marketplaces. in *Proceedings of the 32nd Hawaii International Conference on System Sciences*, (1999).