School of Computing Science,
University of Newcastle upon Tyne

A Model of Trust Derivation from Evidence for Use in Recommendation Systems

Georgios Pitsilis and Lindsay Marshall

Technical Report Series

CS-TR-874

November 2004

Published by the University of Newcastle upon Tyne,
School of Computing Science, Claremont Tower, Claremont Road,
Newcastle upon Tyne, NE1 7RU, UK.

A Model of Trust Derivation from Evidence for Use in Recommendation Systems

Georgios Pitsilis[1] and Lindsay Marshall

School of Computing Science, University of Newcastle Upon-Tyne,
Newcastle Upon Tyne, NE1 7RU, U.K.
{Georgios.Pitsilis,Lindsay.Marshall}@ ncl.ac.uk

Abstract. In this paper we present a technique for modeling trust relationships that can be used in environments where agents take part in a recommendation system. In such an environment, agents collaborate with each other with the common aim of providing accurate recommendations to each other. In the model we use techniques taken from collaborative filtering to express trust properties as beliefs. We also compare our model with existing techniques that can be used to map types of behaviors to trust values. Finally, we describe the requirements of a protocol that could be used for the deployment of such a model in a real distributed environment.

Keywords: *Recommendation Systems, Subjective Logic, Trust Evaluation.*

1 Introduction

Recommender systems have become popular recently as they are widely used in E-commerce online services where they offer suggestions about items customers might also like to buy. Their contribution comes in two forms, either as predicted ratings of services that a user wants to know about, or as lists of services that users might find of interest. The effectiveness of a *Recommender system* can be measured by the accuracy of the predictions that it makes. *Collaborative filtering* (CF) [1] is the most widely known technique used in Recommender systems and is based on the idea of making predictions using similarity metrics to correlate users.

However, *Recommender Systems* and particularly C*ollaborative Filtering* are not perfect and as it is well-known that they seem to have weaknesses such as a low quality of predictions which are known as the *false negative* and *false positive* problems [2], caused by sparsity in the dataset. Also, the architectural characteristics of CF are known to be vulnerable to attacks from malicious and libelous users.

CF systems employ statistical techniques to develop virtual relationships between users. In this way, neighborhoods of users can be formed consisting of those who have a history of agreeing and thus are assumed to be similar.

[1] Scholar of Greek State Scholarships Foundation (IKY)

Trust is a notion which can also express integrity in the relationships between entities. Under the appropriate circumstances with regard to a common purpose, trust relationships can support transitivity [6] whereas similarity generally does not.
The rest of the paper is organized as follows. In the next section, there is a more detailed description of the problem. Section 3 describes related work in the field and in section 4 we analyse our approach to the problem, showing the formulae we introduced, the tests we performed and some comparative results. Finally, in section 5 we discuss some future issues concerning the applicability of our proposed method.

2 Motivation

As we mentioned in the previous paragraph, sparsity in recommendation systems is the main reason for them behaving poorly, because not much evidence can be gathered to support a recommendation. This is usually because users themselves are not willing to invest much time or effort in rating items. In existing CF systems users can only be correlated through their common experiences, so in the presence of limited data they turn out to be unable to make accurate predictions. Our idea is to enhance the neighboring base of users by using trust relationships that could have been developed between them so that it can make it possible to reach other members of the community through them.
For such an idea to become applicable, it requires that somehow users will be able to place trust on their neighbors. In some centralized consumer opinion sites [5] it is a requirement that this trust measure should be provided by the users themselves. This, though, necessitates that users should have developed some instinct in judging things in which case that cannot be assured. Poor judging abilities include the danger of establishing relationships with the wrong counterparts. Our approach to this issue is to introduce a technique for mapping between similarity measures and trust.
In our model we use ordinary measures of similarity taken from *Collaborative Filtering* to form the potential trust between the correlated entities which would be propagated in some identical way to a word-of-mouth scheme. The trust that the first entity should place on the distant one is derived through the trust graph. Finally, by transforming the value back into similarity measure terms, it could then be made appropriate for use in CF schemes.
In our model we express trust in the form of opinions as they are modeled in subjective logic [4]. In this theory *trust* is considered as a subjective measure and introduces the important idea that there is always imperfect knowledge when judging things.
The use of trust in transitive chains requires the existence of a common purpose [6] which needs somehow to be derived from or given by a specific transitive chain. This has either to be modeled from relevant evidence or somehow trustors must be enabled to derive it from past experiences.

3 Background Research

Trust has long been a concern for scientists and much work has been done to formalise it in computing environments [7][8]. An important characteristic is that it is context specific. It is also related to tasks in the sense that entities are trusted to perform a particular task. A simplistic approach would be to determine the levels of trust and distrust that should be placed on some entity from its probabilistic behavior as it is seen from a trustor's point of view. In that sense, trust can be thought of as the level of belief established between two entities in relation to a certain context. In uncertain probability theory [9] the metric which expresses belief is called *opinion*. Because there is always imperfect knowledge as opinions are based on observations, lack of knowledge should be considered when assessing them.

Subjective Logic [4] is a framework for artificial reasoning that deals with the absence of both trust and distrust by introducing the uncertainty property in opinions. *Subjective Logic* uses a simple intuitive representation of uncertain probabilities by using a three dimensional metric that comprises belief (b), disbelief (d) and uncertainty (u). The relationship between b,d and u is expressed as b+d+u=1 which is known as the *Belief Function Additivity Theorem*. Building up opinions requires the existence of evidence. Even though opinions in the form (b,d,u) are better manageable due to the quite flexible calculus that opinion space provides, evidence however is usually available in other forms, that are essentially more understandable by humans. The *Beta Distribution Probability Function* can offer an alternative representation of uncertain probabilities [3], making it possible to approximate opinions from behavioral data. However, data in that evidence space are considered as sets of observations and therefore they must be provided strictly in binary form representing the possible two outcomes of a process, x or $\bar{x}$. So, a behavior is described by the number of x and $\bar{x}$ that derives from the set of observations. In [4] there is a mapping between Evidence Spaces and Opinion Spaces where the uncertainty property (u) is solely dependent on the quantity of observations. In contrast, other similarity based approaches such as that in [10] are based on the idea of getting the users linked together indirectly using predictability measures, but these have not been tested in real environments.

As we mentioned, the requirement for trust to become transitive in long chains requires that a common purpose exists along the chain. According to this, only the last relationship should concern trust about a certain purpose and all the other trust relationships in the chain should regard recommending abilities about the given purpose.

It worth mention the existence of another approach to making recommendation systems trust-enabled [11] in which there is no distinction between functional and recommender trust.

4 Our Approach

In general, trust models are used to enable the parties involved in a trust relationship to know how much reliance to place on each other. Our model aims to provide a

method for estimating how much trust two entities can place in each other, given the similarities between them.
The problem that emerges when Trust is to be used in a recommendation system is the fact that the entities involved usually provide their views in the form of ratings about items and not as their trust estimates about other entities. This means, making the model trust enabled requires that all this info, which so far has been expressed in the form of ratings, must be transformed into trust values. And this of course requires a transformation method.
In order to achieve that, we consider the ratings that users have given to items as the behavioral data required for the composition of their opinions. In our model we assume that the level of trust that develops between every pair of entities is based on how similar they perceive each other's choices to be. We use the Pearson coefficient, as this is the best known and most suitable coefficient for this type of application. This coefficient can take values between -1 and 1 where two entities are considered to as having higher similarity when their Pearson values are close to 1 and as completely dissimilar when the Pearson Coefficient is -1. A value of 0 would mean that there is no relationship between the two entities at all.
Unlike the Beta distribution mapping to Opinions, in our model we describe *Uncertainty* by using both quantitative and qualitative criteria from the evidence.
Similar to the Beta distribution, the rule for applying quantitative criteria obeys the rule that uncertainty should be inversely proportional to the quantity of Evidence. As to the quality of the data, we re-defined the perception of *Uncertainty* as the inability of some entity to make accurate predictions about the choices of the counterpart in the relationship. A low ability value should be the result from the existence of conflicting data and this should make the observer unable to fill in the uncertainty gap. When there are not enough observations to distinguish rating trends data might appear to be highly conflicting.

4.1 Usage

Bearing in mind the idea that those entities whose ratings can be accurately predicted should be considered as trustworthy sources of information, the uncertainty in such relationships should be lower. We propose the following formula to model uncertainty from prediction error:

$$u = \frac{1}{k}\sum_{x=1}^{k}\frac{|p_x - r_x|}{m} \qquad (1)$$

where k is the number of common experiences (ratings) of the two entities that take part in a relation, p_x is the predicted rating of item x calculated using some prediction calculation formula and r_x is the real rate that the entity has given to item x. m represents the maximum value that a rating can take and it is used here as a measure of rating. As can be seen, uncertainty is inversely proportional to the number of experiences. This agrees with the definition of uncertainty we presented in the previous section.
Unlike Beta mapping where u tends to 0 as the number of experiences grow, in our model the trend remains quite uncertain because it is also dependent on the average

prediction error. In the extreme case where there is high controversy in the data, u will reach a value close to 1, leaving a small space for belief and disbelief.
Another interesting characteristic of our model is the asymmetry in the trust relationships produced, which adheres to the natural form of relationships since the levels of trust that two entities place on each other may not be necessarily the same.
As regards the other two properties b (belief) and d (disbelief), we set them up in such a way that they are dependent on the value of the Correlation Coefficient CC. The formulae we use are:

$$b = \frac{(1-u)}{2}(1+CC) \quad (2), \qquad d = \frac{(1-u)}{2}(1-CC) \quad (3)$$

As can be seen, the ratio of belief and disbelief is shaped by the CC value. In this way, a positive Correlation Coefficient would be expected to strengthen the belief property at the expense of disbelief. In the same way, disbelief appears to be stronger than belief between entities that are negatively correlated (CC<0).
These two formulae can be used in the opposite way too, for estimating how similar the two entities should consider each other, given their Trust properties. The asymmetry in the trust relationships is mainly responsible for having unequal similarities in the normal and the opposite relationship. The different points of view are responsible for this difference as well as the formula used to work out the predictions p_x in (1). Formulae proposed in [10] as well as Resnick's [12] empirical formula for the Grouplens CF system can be used for the above purpose.

4.2 Test method

In this section we present experimental results in the form of a comparative study that shows the accuracy of our modeling method. We compare our Evidence to Opinion mapping against a modeling based on Beta distribution. We used a dataset taken from a real CF system known as *MovieLens* and we modeled opinions using both schemata and finally we demonstrate how close they appear to be.
MovieLens [13] is a movie recommendation system based on collaborative filtering established at the University of Minnesota. The whole dataset is publicly available and contains 1,000,209 anonymous ratings of approximately 3,900 movies made by 6,040 users who joined the service over the year 2000.
As stated in section 3, the Beta Distribution Function requires that behavioral data should be expressed in binary form, which refer to the two possible outputs of a process that characterize it as satisfactory or non-satisfactory. This, though, makes the modeling inflexible when ratings are expressed using continuous values or in numerical discrete alternatives. In the case of the data set used for our experiments, the ratings were available in discrete values ranging from 1 to 5. We restricted the test to a subset of the MovieLens database based on 100 users. In total, the testing dataset comprised 12,976 ratings. The analysis we performed on the dataset showed an asymmetric distribution of the ratings with mean=3.61, sd^2=1.24, median M=4 and with a skew to the left. The value 4 for the median can be explained by the fact that people tend to be kind when they rate things they have experienced themselves.

We faced two challenges when carrying out this experiment. First, how to make the experimental dataset suitable for representing evidence for the Beta distribution, and second what measures to use for the comparison.
Because there was no data in an appropriate form supplied by users showing how much they trust each other lead us to generate the weights that should be placed on their relationships artificially. Beta modeling requires that evidence should be provided in binary form, x or $\bar{x}$ (meaning satisfactory or unsatisfactory) to represent how an entity would perceive the behavior of another party.
We defined our own criterion for judging a behavior for how every single item was rated. Let us call $R_{A,k}$ the rate that user A gave to item k and $R_{B,k}$ the rate of user B to the same item. A relatively long distance between $R_{A,k}$ and $R_{B,k}$ should be considered –subjectively judged – by A or B as unsatisfactory behavior of the other counterpart. As can be seen, such a rule requires a criterion for judging a behavior as x or $\bar{x}$.
In our experiment we used the Median value as the barrier for characterising a behaviour as bad if the two ratings have been placed on different sides. For example a case where $\mathbf{R_{B,k}}$=3 and $\mathbf{R_{A,k}}$=5 should be taken as $\bar{x}$. The median reflects the way that users rate items. Finally, we choose one of the four possible scenarios of the table that characterise a behaviour.

	$\mathbf{R_{A,k}}$>M	$\mathbf{R_{A,k}}$<M
$\mathbf{R_{B,k}}$>M	x	$\bar{x}$
$\mathbf{R_{B,k}}$<M	$\bar{x}$	x

Figure 1. Truth table of Evidence

Once all the pairs of common ratings have been examined, we transform the evidence to opinions (b,d,u) as described in [4].

4.3 Comparative Results

In our test, 8782 trust relationships were tested from a sample dataset of 100 users, and the results are given below.
In order to be able to compare the opinions created by each model, we converted them to a plain probabilistic value which by convention called *Probability Expectation* (PE). The PE can be interpreted as saying that the relative frequency of both counterparts in the relationships agreeing in taste is somewhat uncertain and the most likely value is E(x)=b+au. A formal definition of PE can be found in [4]. In the experiment we measured how close the two derived opinions are by comparing their probability expectations. Therefore, the values shown in our results are in terms of this measure.
In figure 2 we present the divergence between our modeling and the Beta distribution function, the measurements being derived from the relative difference (%) between the two probability expectations. The results have been grouped for various classes of common experiences that constitute an opinion, to show how the number of experiences affects the distance. The second column indicates how many relationships from the dataset have been found to have a number of experiences that belongs to that class.

Class	num. of Common Experiences	sample size	mean (%)	sd (%)
1	[2-3)	624	17.50	8.64
2	[3-5)	1326	16.67	8.17
3	[5-10)	2277	13.77	7.67
4	[10-20)	2191	11.29	7.03
5	[20-40)	1417	10.10	6.88
6	[40-60)	471	11.22	7.33
7	[60-80)	195	11.98	7.75
8	[80-100)	101	12.90	7.68
9	[100-150)	115	10.64	7.71
10	[150-200)	46	10.52	8.49
11	[200-250]	10	11.26	7.81

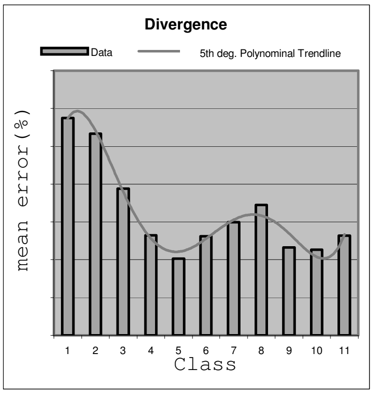

Figure.2: *mean* and *std.dev.* of divergence for various numbers of common experiences

Figure.2 shows increased divergence between the two modeling techniques for small numbers of common experiences and it can be explained as the result of the noisy behavior of the clustering coefficient. In these categories the quality of predictions is quite uncertain.
The small increasing trend in divergence that is noticeable as the numbers of common experiences grows (class 4) is due to the poor data set used in those categories since only a very small sample size existed in this class of common experiences.

4.4 Discussion of the results

From the results it can be seen that the method converges to Beta modeling when the agents have at least 20 common experiences and having more does not provide any extra benefit. It can also be seen that the method performs better when the considered opinions have been built upon at least 20 common experiences at which point the divergence stabilizes.
Even though both methods give slightly different results (differ by around 10%), there is no real situation which could be used as a point of reference to evaluate how accurate, in absolute measures, each method is.

5 Future Work

We intend to apply this technique to a real recommendation system, with the expectation that it will improve the quality of the derived recommendations. Another idea is to make use of the web-of-trust that would evolve from the establishment of direct trust relationships between users. Our aim is to improve the recommendations by exploiting the experiences of any entities not neighbouring the querying one but which can be reached via the web-of-trust.

The question that arises from this is how accurate these predictions can be. Short tests we performed, showed a significant increase in the coverage, which translates into reduced sparsity, without significant impact on the error in predictions. Our short-term plans include a thorough study and analysis of the various parameters that impact the results as well as a performance analysis of the resulting system. The long-term plans include the deployment of such a solution in a totally distributed recommendation system.

For the above scenario that incorporates graphs of opinions, a querying entity is required to know, not only the *direct* trust value for its neighbors, but also how good they are in recommending other entities. In other words, how much it trusts their recommendation abilities, in respect of the given purpose. This is what is called *recommendation* trust. A *common purpose* along a trust chain must exist because this is what makes the trust transitive.

Recommendation trust can be derived in a similar way to that described for direct trust in this paper. The basic idea is that someone's (lets call it the trustee) recommendation trust can be estimated by some other entity (lets call it the trustor) by comparing any recommendations that the trustee has provided in the past about things for which trustor also maintains its own evidence. Then the trustor, by comparing its relevant personal experiences with the trustee's recommendations, will be able to estimate how good in doing recommendations the trustee has been. Similar to *direct* trust, *recommendation* trust is a subjective measure, which means, every trustor has to maintain its own picture of its environment.

The ad-hoc way we chose to code the positive and negative evidence for the Beta distribution necessitates more tests against other alternative coding techniques and use of different statistical measures (e.g the Mode value instead of Median).

No matter how successful recommendations such an architecture can provide, there are weaknesses concerning security for the recommendation systems that must also be tolerated. In particular, any deployed solution must be resistant to attacks from users that try maliciously to influence the system. In the case of deploying the solution in a distributed recommendation system, the communicated experiences during the trust calculation must be done through some secure protocol.

6 Conclusion

We presented an empirical technique for modeling the trustworthiness of entities using evidence that describes their rating behavior. The novelty comes from the shaping of the derived uncertainty which is dependent on a predictability measure and thus on the value of the Evidence. We coded our derived trust opinions into metrics taken from Shaferian belief theory and we attempted an evaluation of our model against an identical one which uses the Beta distribution function for mapping evidence to opinions. From the evaluations it appears that both methods produce very similar results.

The strong points of the proposed technique can be summarised as its ability to incorporate similarity measures in its properties, its use of qualitative as well as quantitative measures to derive opinions and its flexibility in accepting datasets of continuous values rather than binary, which makes the method suitable for CF recommendation systems.

7 References

[1] D.Goldberg - D.Nichols - B.M.Oki - D.Terry, "Using Collaborative filtering to weave an information tapestry", Communication of ACM, 35(12):61-70, 1992

[2] B.Sarwar – G.Karypis – J.Konstan – J.Reidl, "Analysis of Recommendation Algorithms for E-Commerce", In Proceedings of the Second ACM Conference on Electronic Commerce pg 158-168,ACM Press 2000.

[3] A. Jøsang and R. Ismail. *The Beta Reputation System*. In the proceedings of the 15th Bled Conference on Electronic Commerce, Bled, Slovenia, 17-19 June 2002.

[4] A. Jøsang, "A Logic for Uncertain probabilities", International Journal of Uncertainty, fuzziness and Knowledge based systems, Vol.9,No.3, June 2001.

[5] http://www.epinions.com

[6] A. Jøsang, E. Gray and M. Kinateder. *Analysing Topologies of Transitive Trust*. In the proceedings of the Workshop of Formal Aspects of Security and Trust (FAST) 2003, Pisa, September 2003.

[7] S.Marsh, "Formalizing Trust as Computational concept", PhD Thesis, University of Stirling, Scotland 1994.

[8] A.Rahman – S.Heiles,"Supporting trust in Virtual Communities", In proceedings of International conference On System Sciences, Jan 4-7-2000, Hawaii.

[9] G.Shafer, "A Mathematical Theory of Evidence", Princeton University Press. 1976

[10] Aggarwal, C.C. Wolf, - J.L. - Wu K. - Yu, P.S. (1999). Horting Hatches an Egg: A New Graph-theoretic Approach to Collaborative Filtering. In Proceedings of the ACM KDD'99 Conference. San Diego, CA, pp.201-212.

[11] P.Massa – P.Avesani, "Trust-aware Collaborative Filtering for recommender Systems", CoopIS/DOA/ODBASE (1) 2004: 492-508

[12] P.Resnick - N.Iacovou - M.Suchak - P.Bergstrom - J.Riedl, "Grouplens. An OpenArchitecture for Collaborative filtering of Netnews, from Proceedings of ACM 1994, Conf. On Computer Supported Cooperative.

[13] Bradley N. Miller, Istvan Albert, Shyong K. Lam, Joseph A. Konstan, John Riedl. (2003). MovieLens Unplugged: Experiences with an Occasionally Connected Recommender System. InProceedings of ACM 2003 International Conference on Intelligent User Interfaces (IUI'03) (Accepted Poster), January 2003.