Electronic Commerce Research and Applications 8 (2009) 37–47

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

journal homepage: www.elsevier.com/locate/ecra

Fuzzy computational models for trust and reputation systems

Kamal K. Bharadwaj, Mohammad Yahya H. Al-Shamri *

School of Computer and Systems Sciences, Jawaharlal Nehru University, New Mehrauli Road, New Delhi 110067, India

ARTICLE INFO

Article history:
Received 17 December 2007
Received in revised form 11 June 2008
Accepted 7 August 2008
Available online 14 August 2008

Keywords:
Trust
Reputation
Reciprocity
Recommender systems
Collaborative filtering
Fuzzy computational models

ABSTRACT

In the recent past, a considerable research has been devoted to trust and reputation mechanisms to simplify complex transactions for open environments in social networking, e-commerce, and recommender systems (RS). In real life, we come to know about others through our social circle according to their reputation which is a public view. However, it is not always adequate to depend solely on the public view and therefore a trust measure is required to give a personalized view of the future encounters with a specific partner. In this paper, we propose fuzzy computational models for both trust and reputation concepts. Reciprocity and experience are used for trust modeling while the proposed reputation model is a fuzzy extension of beta reputation model. A two-level filtering methodology is proposed to benefit to a large extent from both the concepts separately. In order to justify the proposed models, we compared them with the existing reputation models for movie RS. The experimental results show that the incorporation of trust and reputation concepts into RS indeed improves the recommendation accuracy and establish that our models are better than beta and the popular eBay reputation models.

1. Introduction

The emerging online environment let people buy and sell goods, play, and recommend products for each other without knowing the remote partner. All of these encounters are only possible if partners trust each other. Consequently they can delegate tasks and decisions to an appropriate person and therefore improving the quality of online markets [1,2]. However, insufficient trust in e-commerce could lead to users '*staying away*' from the technology altogether. Despite obvious importance of trust and reputation, transferring them to computer science is still a challenging problem. Computer scientists from many areas, e.g., security, semantic web, and e-commerce, are still working on the transfer of these concepts to be used in computer mediated transactions. It is often hard to assess the trustworthiness of remote partners in the computer mediated transactions and processes because familiar styles of interaction are still far from being computationally modeled easily. Actually, physical transactions and traditional forms of communication like body language, gestures, and facial expressions allow people to assess a much wider range of cues related to trustworthiness than the current computer mediated communication can do.

Amongst different areas of computer science, e-commerce seems to be the most benefitted area from the trust and reputation concepts. Recently, very big electronic communities have been established and grown very fast. Unfortunately, a major weakness of e-markets is the raised level of risk associated with the loss of the notions of trust and reputation. The consumer is usually forced to accept the *'risk of prior performance'*, i.e. to pay for services and goods before receiving them, which can leave him in a vulnerable position. On the other hand, the service provider has much more information about the product quality than the customer, as long as he will receive payment before shipment in most cases. The research work in the recent past [1,3–8] shows that this information asymmetry can be mitigated through trust and reputation concepts. The idea is that even if the consumer cannot try the service in advance, he can be confident that it will be what he expects as long as he trusts the service provider.

Various definitions have been given in the literature to both trust and reputation concepts. For example, Ruohomaa and Kutvonen [7] defined trust as the extent to which one partner is willing to participate in a given action with another partner, considering the risks and incentives involved. On the other hand, the reputation is defined by Mui et al. [8] as a perception a partner creates through past actions about his intentions and norms. Although the way of incorporating trust and reputation concepts in e-commerce varies from system to system but they provide an incentive for good behavior and a punishment for bad behavior. However, trust and reputation concepts are social fuzzy concepts thus fuzzy computational models for both the concepts would reflect their actual value for enhancing the system accuracy.

The rest of this paper is organized as follow: trust and reputation concepts in literature are given in Section 2. The proposed

* Corresponding author. Tel.: +91 11 9810196636 (mobile); fax: +91 11 26717528.

E-mail addresses: kbharadwaj@gmail.com (K.K. Bharadwaj), mohamad.alshamri@gmail.com (M. Y. H. Al-Shamri).

1567-4223/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.elerap.2008.08.001

fuzzy computational models for trust and reputation are presented in Section 3 and are employed for movie RS in Section 4. Computational experiments and results are given in Section 5. Finally, the last section concludes the paper and gives some future research directions.

2. Trust and reputation in literature

Trust and reputation concepts get much attention due to the need for reliable automatic tools for Web services. The traditional cues of trust and reputation of physical encounters are missing in Web interactions. As a consequence, electronic substitutes are required for these concepts so that the quality of online services can improve especially for e-commerce and entertainment services. However, trust and reputation are quite challenging to define as they manifest themselves in many different forms. The literature on trust and reputation is also quite confusing because these terms are being used with a variety of meanings [2,3,5,7–10]. As indicated by the recent survey of the developments in this area [1], there is considerable confusion around the terminology used to describe trust and reputation concepts, and therefore a consistent terminology is required. However, the following properties [2] are regularly assigned to trust, and are relevant when transferring it to computer science.

- Trust is subjective and therefore asymmetric. The levels of trust that two partners place on each other may not be necessarily the same.
- It is context dependent.
- It is dynamic, meaning it can increase with positive experience and decrease with negative experience. Thus trust is non-monotonic.

There is a conflict about trust transitivity in the literature that arises from the way of treating trust as a global or a subjective property. For those who treat trust as a global property it is transitive while it is not for those who treat it as a subjective property. For example, Marsh [10] pointed out that trust is not transitive. It is not transitive also as a global property over arbitrary long chains, since this will end in conflict regarding distrust. The research in this field report three main stages, namely modeling, management and decision making [2]. McKnight and Chervany [9] categorize trust into personal trust, impersonal trust, and dispositional trust. Personal trust describes trust between people while impersonal trust is not bound to a person but arises from social or organizational situation. Dispositional trust is the person's general attitude towards the world. This paper focuses on modeling of reputation as a global property and personal trust as a subjective property and therefore trust is not transitive. Moreover, trust and reputation are classified according to the information sources that they take into account for computing trust and reputation values. Accordingly, in the following we will touch three main points, namely trust and reputation definitions, information sources, and the previously proposed models.

2.1. Reputation and trust definitions and information sources

Different disciplines handle trust differently according to their own perceptions and what fits their specific goals. Accordingly, the existing trust definitions span different disciplines like, Psychology, Sociology, Philosophy, Marketing, E-Commerce, and Computer Science [5]. For example Gambetta [11] stated that 'trust is the subjective probability by which an individual *A* expect that another individual *B* performs a given action on which its welfare depends'. On the other hand, Ostrom's work [12] stated that 'Reputation is the perception that an agent creates through past actions about its intentions and norms'. Sometimes the two concepts are used interchangeably. However, recent researchers have pointed out that there is a clear difference between them without ignoring that there is some correlation between the two concepts in some cases [1,3]. This is in agreement with the ground basics of these concepts in sociology. Accordingly, two distinct definitions are tried out for trust and reputation concepts.

Two main sources of information [1,3] are frequently used for trust and reputation modeling, namely *direct information* and *word of mouth information* (witness information). The direct information is the experience based on the direct interactions with a specific partner while witness information is the information comes from other members of the community. We believe that trust and reciprocity are one-to-one relationship (personal relation) while reputation is a many-to-one relationship (public relation). Therefore, the direct information is suitable for trust and reciprocity modeling while the witness information is appropriate for reputation modeling to reflect the community's opinion in general. This agrees with almost all recent literature which treats reputation as a global property shared by all the observers but they are varying in reflecting this opinion in their models. In this case, the reputation score reflects the past interactions of the community with the individual being evaluated. This score is globally available to all members of the community and updated each time a member issues a new evaluation of an individual. On the other hand, trust is a personal and subjective property assessed particularly by each individual. That means you cannot trust anybody before you get a direct contact with him. There are special cases when the witness information is used but this is when there is a lack of direct information and that should be related to reputation not trust. In this direction, we have to talk about the trust of a partner a_x from the point of view of *a* partner a_y.

2.2. Previous reputation and trust models

Many computational trust and reputation models have appeared [4,5,8,10,13–16] in the last few years. In the computer science literature, Marsh [10] is among the first to introduce a computational model for trust in the distributed artificial intelligence community but he did not model reputation in his work. As he has pointed out, several limitations exist for his simple trust model. Marsh model takes into account direct information and differentiates three types of trust, namely basic trust, general trust, and situational trust. Moreover, Marsh introduces the notion of "reciprocation" as a modifier for trust measures. Actually, the previous models vary according to the information sources and the constituting properties of each concept. Some authors treat reputation as a property of trust or vice versa while others not. For example some authors treats reputation as a means to build and update trust after a certain number of successful transactions [7,15] while Mui et al. [8] assumed that trust, reputation and reciprocity reinforce each other. However, trust and reciprocity are personal properties and therefore can reinforce each other while reputation is a global property that can be overridden by personal experience which is the main source for trust and reciprocity modeling. Some people could say *'we trust you because of your good reputation'* while others could say *'we trust you despite your bad reputation'*. In the later case, the personal experience and intimate relations override reputation [1]. This variability of opinions cannot be captured if we model reputation based on trust or vice versa. By building separate models according to separate sources of information, we can maintain the validity of both cases, reinforcement between the two concepts if both values are high or trust overrides reputation if the trust measure is high. The unclear separation between both concepts leads Zacharia [13] to design Histos system to overcome the lack of personalization that Sporas system has about reputation scores. Histos system takes into account both direct and witness information sources.

In this paper we concentrate on centralized reputation systems where the reputation centre collects ratings from all community members. The reputation scores are computed according to many ways and are regularly updated as a function of the received ratings. For example, eBay – the popular e-commerce site with a community of over 50 millions registered users (eBay.com) – used a simple summation such that the running total reputation score of a given partner is the sum of positive ratings minus the sum of negative ratings. Other alternatives to simple summation are average of ratings (Amazon.com) and weighted average to put more emphasis on the most recent transactions or to highlight some factors more than others. A more complex system is the beta reputation system [16] which provides a theoretically sound basis for computing reputation scores. The beta family of distributions is a continuous family of distribution functions indexed by two parameters α and β. The beta probability distribution function denoted by beta($p|\alpha,\beta$) can be expressed using the gamma function, Γ, as below [16,17]:

$$\text{beta}(p|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \tag{1}$$

where $0 \leqslant p \leqslant 1, \alpha > 0, \beta > 0$, $p \neq 0$ if $\alpha < 1$, and $p \neq 1$ if $\beta < 1$. The probability expectation value of the beta distribution is:

$$E(p) = \frac{\alpha}{\alpha+\beta} \tag{2}$$

If the number of positive observations is N^+ and the number of negative observations is N^- then $\alpha = N^+ + 1$ and $\beta = N^- + 1$ As a consequence, the beta reputation function becomes:

$$f(p|N^+, N^-) = \frac{\Gamma(N^+ + N^- + 2)}{\Gamma(N^+ + 1)\Gamma(N^- + 1)} p^{N^+}(1-p)^{N^-} \tag{3}$$

Based on the most natural way to define the reputation score as a function of the expectation value, Jøsang and Ismail [16] defined reputation score by the following:

$$\text{rep}(a_i) = 2(E(f(p|N^+, N^-)) - 0.5) = \frac{N^+ - N^-}{N^+ + N^- + 2} \tag{4}$$

This can be interpreted as saying that the relative frequency of a positive outcome in the future is somewhat uncertain, and the most likely value is $E(f(p|N^+, N^-))$. Formula (4) will be used later for deriving the reputation scores between partners. To take into account multiple sources G, N^+ and N^- are computed as [16]:

$$N^+ = \sum_{j\in G} N_j^+ \quad \text{and} \quad N^- = \sum_{j\in G} N_j^- \tag{5}$$

where N_j^+ and N_j^- are the positive and negative observations for source j, respectively.

3. The proposed fuzzy computational models for trust and reputation concepts

The basic idea of trust and reputation systems is to let partners rate each other after the completion of an encounter. The system then aggregates these ratings for a given partner to derive a personalized trust measure or a global reputation score [1]. However, either trust or reputation are usually used as crisp attributes. That does not effectively reflect the social meanings of these two concepts where most of human perceptions are fuzzy. Our attempt in the following section is to discuss definitions for trust, reciprocity and reputation concepts based on their social meanings and accordingly propose fuzzy computational models for them.

3.1. Trust, reputation and reciprocity definitions

Despite obvious usefulness of trust and reputation, conceptual gaps exist in current models about them. In real life we come to know about others through our social circle through their reputation. However, to what extent the piece of information obtained via third party is correct and as a result there is uncertainty in the recommendation. This motivates us to differentiate between trust and reputation concepts and use a consistent terminology to define both concepts according to their social meaning and usage as close as possible. Accordingly fuzzy computational models are built based on these definitions. Firstly, the appropriate definition for trust is [8]:

Definition 1 (*Trust*). Trust is a subjective expectation a partner has about another's future behavior based on the history of their encounters.

Obviously, this definition treats the trust as a subjective property computed based on the two partners concerned in a dyadic encounter. The term '*subjective expectation*' is used to emphasize that trust is a personalized summary that an individual has toward another based on a number of former encounters between them. Secondly, the concept of reciprocity is closely related to that of trustworthiness. The social dictum "*Be nice to others who are nice to you*" seems to be well permeated in our society for encouraging social reciprocation. This social dictum reflects exactly the concept of *reciprocity* as studied by evolutionary biologists. This paper uses the following definition for reciprocity [8]:

Definition 2 (*Reciprocity*). Reciprocity is the mutual exchange of deeds (such as favor or revenge).

Thus the reciprocity is a symmetric property while the trust is an asymmetric property. This indicates that there is another property that should introduce asymmetry into the trust measure if we treat reciprocity as the symmetric part of trust. Finally, the Concise Oxford dictionary gives a definition for reputation [1] which is in agreement with the social network researchers' point of view.

Definition 3 (*Reputation*). Reputation is what is generally said or believed about a person's or thing's character or standing.

Thus reputation exists only in a community which is observing its members in one way or the other. Accordingly, reputation is the collected and processed information about one partner's former behavior as experienced by others. The evolved score is the aggregation and is globally visible to all members of the network. Thus the reputation score reflects a global view of each individual and therefore suffers from a lack of personalization, unlike trust measure. Something that is bad for me could be acceptable for others and the other way around [3]. By including both concepts separately in our work we maintain the advantages of both concepts. Thus, high trust measure could override even high reputation scores based on the past direct experience with the others.

3.2. The proposed fuzzy computational models for trust and reputation

Trust and reputation systems are rating systems where each individual is asked to give his opinion after completion of each encounter in the form of ratings. More formally, let $A = \{a_1, a_2, \ldots, a_M\}$ be the set of all partners (users), where M is the number of partners in the system. We assume each partner will rate the other after completing the encounter. An encounter $e_k \in E$ is an ordered pair, $e_k(a_i, a_j) = \langle r_{a_i}^{e_k}(a_j), r_{a_j}^{e_k}(a_i)\rangle$, where $r_{a_j}^{e_k}(a_j)$ is the rating partner a_i has given to partner a_j for an encounter e_k. The rating scale Z can take different forms, for example $Z_{\text{eBay}} = \{-1, 0, +1\}, Z_{\text{beta}} = \{-1, +1\}$, and $Z_{\text{MovieLens}} = \{1, 2, 3, 4, 5\}$.

Binary ratings are considered insufficient to capture various degrees of judgment, therefore Z should give more than one choice for positive/negative ratings besides a neutral rating. As a consequence, ratings can be easily assigned and understood by human users and therefore a more accurate judgment can be obtained [2]. For example, $Z_B = \{-2, -1, 0, +1, +2\}$ can be translated into discrete form as very bad, bad, average, good, and very good, respectively. The set of ratings partner a_i has given to partner a_j is $S_{a_i}(a_j) = \{r^{e_k}_{a_i}(a_j) | e_k \in E\}$ and the whole past history of partner a_i is $H_i = \{S_{a_i}(a_j) | \forall a_j (\neq a_i) \in A\}$. It is to be noted that usually $|S_{a_i}(a_j)| = |S_{a_j}(a_i)|$ and when both partners do not rate each other, i.e. if a_i has no encounters at all with a_j and vice versa then $S_{a_i}(a_j) = S_{a_j}(a_i) = \emptyset$, where $\emptyset$ denotes an empty set.

3.2.1. Fuzzy computational model for trust

Trust is a complex concept that constitutes many properties. Amongst these properties one is symmetric whereas the other is asymmetric. The most appropriate property to define the symmetric part is the reciprocity while the partner's experience defines the asymmetric part. The reciprocity is the mutual favor or revenge and therefore to model it, we need to find the agreement (both individuals are satisfied or unsatisfied) and disagreement (only one of them is unsatisfied) between two partners. To do so, we can define two fuzzy subsets on each partner's ratings (universe of discourse), namely satisfied and unsatisfied. The membership values of satisfied and unsatisfied fuzzy subsets for a given encounter always sum up to one, for example, 70% satisfaction indicates 30% unsatisfaction. From these two fuzzy sets we can find the agreement and disagreement between the two partners. The satisfied and unsatisfied fuzzy subsets for partner a_i are defined as below:

$$\text{satisfied}(a_i) = \{\text{sat}_{a_i}(e_k) | e_k \in H_i\} \tag{6}$$

$$\text{unsatisfied}(a_i) = \{\text{unsat}_{a_i}(e_k) | e_k \in H_i\} \tag{7}$$

where $\text{sat}_{a_i}(e_k)$ and $\text{unsat}_{a_i}(e_k)$ are membership values of a_i's ratings for e_k in the fuzzy subsets satisfied(a_i) and unsatisfied(a_i), respectively. Fig. 1 gives a simple triangle membership functions for satisfied(a_i) and unsatisfied(a_i) fuzzy subsets that are given by the following:

$$\text{sat}_{a_i}(e_k) = \begin{cases} 0 & r^{e_k}_{a_i}(a_j) = z_* \\ \frac{(r^{e_k}_{a_i}(a_j) - z_*)}{z^* - z_*} & z_* < r^{e_k}_{a_i}(a_j) < z^* \\ 1 & r^{e_k}_{a_i}(a_j) = z^* \end{cases} \tag{8}$$

$$\text{unsat}_{a_i}(e_k) = 1 - \text{sat}_{a_i}(e_k) \tag{9}$$

where z_* and z^* are the minimum and the maximum possible ratings for a given system with $Z_B = \{z_*, \ldots, 0, \ldots, z^*\}$.

The possible combinations of satisfied and unsatisfied fuzzy subsets define four values for any two partners, namely satisfied–satisfied SS(a_i, a_j), unsatisfied–unsatisfied UU(a_i, a_j), satisfied–unsatisfied SU(a_i, a_j), and unsatisfied–satisfied US(a_i, a_j).

$$\text{SS}(a_i, a_j) = \frac{|\text{satisfied}(a_i) \cap \text{satisfied}(a_j)|}{|\text{satisfied}(a_i) \cup \text{satisfied}(a_j)|} \tag{10}$$

$$\text{UU}(a_i, a_j) = \frac{|\text{unsatisfied}(a_i) \cap \text{unsatisfied}(a_j)|}{|\text{unsatisfied}(a_i) \cup \text{unsatisfied}(a_j)|} \tag{11}$$

$$\text{SU}(a_i, a_j) = \frac{|\text{satisfied}(a_i) \cap \text{unsatisfied}(a_j)|}{|\text{satisfied}(a_i) \cup \text{unsatisfied}(a_j)|} \tag{12}$$

$$\text{US}(a_i, a_j) = \frac{|\text{unsatisfied}(a_i) \cap \text{satisfied}(a_j)|}{|\text{unsatisfied}(a_i) \cup \text{satisfied}(a_j)|} \tag{13}$$

Fuzzy sets literature describes many alternatives for union and intersection of crisp sets. The popular one is min for intersection and max for union. The *min–max* alternative and the definition of fuzzy set's cardinality as given by Zadeh in [18] yield the following:

$$|\text{satisfied}(a_i) \cap \text{satisfied}(a_j)| = \sum_{e_k \in H_i \cap H_j} \min(\text{sat}_{a_i}(e_k), \text{sat}_{a_j}(e_k)) \tag{14}$$

$$|\text{satisfied}(a_i) \cup \text{satisfied}(a_j)| = \sum_{e_k \in H_i \cap H_j} \max(\text{sat}_{a_i}(e_k), \text{sat}_{a_j}(e_k)) \tag{15}$$

There are many possible ways to compute the agreement and disagreement values based on Formulae (10)–(13). However, this paper uses the following definition:

Definition 4 (*Agreement and Disagreement*). For a rating trust system, the agreement and disagreement values between any two partners a_i and a_j are given by the following formulae:

$$\text{agr}(a_i, a_j) = \frac{\text{SS}(a_i, a_j) + \text{UU}(a_i, a_j)}{2} \tag{16}$$

$$\text{disagr}(a_i, a_j) = \frac{\text{SU}(a_i, a_j) + \text{US}(a_i, a_j)}{2} \tag{17}$$

Now we can express reciprocity mathematically in terms of agr(a_i, a_j) and disagr(a_i, a_j) as follows:

$$\text{rec}(a_i, a_j) = (1 - \text{disagr}(a_i, a_j))\text{agr}(a_i, a_j) \tag{18}$$

The mutual exchange of deeds is represented by agr(a_i, a_j) value (both partners are satisfied or unsatisfied). On the other hand, disagr(a_i, a_j) value represents the conflict both partners have. The reciprocity becomes high if the disagreement is low and vice versa. Accordingly, Formula (18) gives disagreement the priority, i.e. when disagr(a_i, a_j) becomes one, the reciprocity value is zero whatever be the agr(a_i, a_j) value. Moreover, Formula (18) depends only on the degree of agreement and disagreement between the two partners to judge reciprocity. However, how reliable is the computed value? A reliability measure (to be discussed later) should be introduced to weight the reciprocity value. A more accurate formula for reciprocity would be:

$$\text{recip}(a_i, a_j) = \text{reliab}(a_i, a_j) \times \text{rec}(a_i, a_j) \tag{19}$$

The asymmetric part of trust could be the experience of the other partner and the confidence you have on that experience. The robustness of rating systems relies on the number of encounters the partner has made in the past which is in turn defines the partner's experience. However, in social life you cannot trust anyone more than yourself if he has less experience than you. Mathematically this can be expressed by multiplying experience value by a confidence factor. In our model the confidence factor partner a_i has in partner a_j and the experience of partner a_j are:

$$\text{conf}_{a_i}(a_j) = \frac{n_j}{\max(n_i, n_j)} \tag{20}$$

$$\text{ex}(a_j) = \frac{n_j}{\max(n_1, \ldots, n_M)} \tag{21}$$

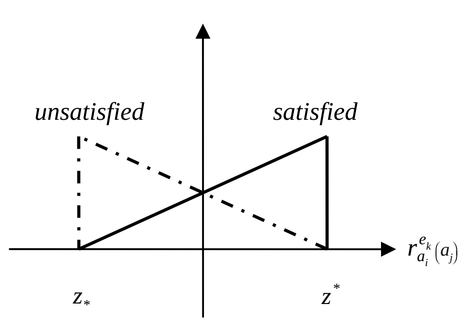

Fig. 1. Membership functions of satisfied and unsatisfied fuzzy subsets.

where $n_i = |H_i|$, the cardinality of H_i and M is the number of partners in the system. The experience value of a_j as seen by a_i is:

$$\text{Exper}_{a_i}(a_j) = \text{conf}_{a_i}(a_j) \times \text{ex}(a_j) \tag{22}$$

Now, trust as a function of reciprocity, experience, and confidence is:

$$\text{trust}_{a_i}(a_j) = \frac{2 \times \text{Exper}_{a_i}(a_j) \times \text{recip}(a_i, a_j)}{\text{Exper}_{a_i}(a_j) + \text{recip}(a_i, a_j)} \tag{23}$$

It is to be noted that reciprocity is a symmetric property while experience is asymmetric and therefore the resulting trust measure will be asymmetric. As you can see, the past history is the basis for trust and therefore you cannot trust anybody unless you have had interaction with him before, but you can rely on his reputation as an alternative to the lack of personal experience. This is exactly what we target by separating trust and reputation concepts. Throughout this paper we used the harmonic mean to aggregate some values because it is robust to large differences between the inputs so that a high value will only be possible if both inputs are high.

3.2.2. Fuzzy computational model for reputation

Jøsang et al. [1] pointed out that there seems to be a lack of coherence in the existing work on trust and reputation models. Authors often propose new systems from scratch without trying to extend or enhance previous proposals. Amongst the existing reputation models, beta reputation model provides a mathematically sound basis for computing reputation scores. This model uses the beta probability density function to represent probability distributions of binary events (positive and negative encounters). However, two modifications could be possible. Firstly, there is no clear differentiation between a single individual's opinion and a group of individuals' opinion. The way of computing the reputation score for a target entity is the same except that the summation of positive and negative encounters for the whole group is used in the later case instead of individual numbers in the former case (see Formula (5)). We think that a good reputation model should differentiate clearly between the way of computing a reputation score for a target entity from an individual and the overall reputation score from the community. According to Definition 3, the whole community where the partner resides has the final word of how much reputation he deserves. Therefore, to compute the reputation for a given partner, each member of the community has to give him a reputation score based on their previous encounters and this score from an individual should be independent of the others opinions. The overall reputation score is the aggregation of all individuals' scores. The second point is that the individual reputation score needs to be associated with a reliability measure. This will measure how reliable is the score received from a given individual. Thus to make beta reputation model more accurate, the following points need to be taken into account.

- The rating scale should give a flexible range of choices (for example, Z_B).
- The beta reputation model is appropriate to compute the reputation scores from individuals while Ordered Weighted Averaging (OWA) [19] operator guided by an appropriate fuzzy linguistic quantifier is the appropriate choice to aggregate all individuals' scores to capture the fuzzy nature of the social aggregation process.
- Each individual reputation score should be associated with a reliability measure.

The eBay and beta reputation models assume binary ratings. However, the system granularity is achieved if a range of ratings is used to reflect a wide range of preferences and judgment. The system can simply use a numerical scale for ratings that can be mapped easily into binary ratings as required for the beta reputation model. Let us assume that the ratings scale is divided equally around the neutral rating (for example Z_B scale), i.e. for each scale point in the positive side, there is a corresponding scale point in the negative side. Thus each pair of points represents binary ratings and therefore we can find the reputation score according to beta reputation model over all the pairs. For clarity let d-pair = $\{-d, +d\}$ represents the partner ratings $r^{e_k}_{a_i}(a_j) = -d$ and $r^{e_k}_{a_i}(a_j) = d$ (for Z_B scale, there are two d-pairs, $\{-1, 1\}$ and $\{-2, 2\}$). Hence a modified version of Formula (4) to compute the reputation score partner a_i has given to partner a_j, $R_{a_i}(a_j)$, is:

$$R_{a_i}(a_j) = \frac{1}{|D|} \sum_{d \in D} \frac{N^{+d}_{i,j} - N^{-d}_{i,j}}{N^{+d}_{i,j} + N^{-d}_{i,j} + 2} \tag{24}$$

where D is the set of all d-pairs, $N^{+d}_{i,j}$ and $N^{-d}_{i,j}$ are the number of positive and negative d ratings partner a_i has given to partner a_j, respectively.

$$\begin{aligned} N^{+d}_{i,j} &= |\{r^{e_k}_{a_i}(a_j) = +d | e_k \in S_{a_i}(a_j)\}| \\ N^{-d}_{i,j} &= |\{r^{e_k}_{a_i}(a_j) = -d | e_k \in S_{a_i}(a_j)\}| \end{aligned} \tag{25}$$

The reputation score given by Formula (24) is based mainly on the number of positive and negative encounters. However, there are a number of neutral encounters. This gives rise to a question like how reliable is the computed score? A reliability measure needs to be introduced to make the derived score more reliable (it will be discussed later). Thus the final reputation score from individuals is:

$$\text{rep}_{a_i}(a_j) = \text{reliab}(a_j) \times R_{a_i}(a_j) \tag{26}$$

The overall reputation score of partner a_j as seen by the whole community is:

$$\text{rep}(a_j) = \text{aggr}(\text{rep}_{a_1}(a_j), \ldots, \text{rep}_{a_{j-1}}(a_j), \text{rep}_{a_{j+1}}(a_j), \ldots, \text{rep}_{a_M}(a_j)) \tag{27}$$

The definition of reputation states "*generally said or believed*" to describe the partner's standing in the community. This can be interpreted in fuzzy logic by using an appropriate fuzzy linguistic quantifier to guide the aggregation process of reputation scores coming from the community members. OWA operators were originally introduced by Yager [19] to provide a means for aggregating scores associated with the satisfaction of multiple criteria, which unifies in one operator the conjunctive and disjunctive behavior:

$$OWA(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} w_j \times x_{\sigma(j)} \tag{28}$$

where σ is a permutation that orders the elements under consideration: $x_{\sigma(1)} \leqslant x_{\sigma(2)} \leqslant \cdots \leqslant x_{\sigma(n)}$. The weights are all non-negative ($w_j \geqslant 0$) and their sum equals one, i.e. $\sum_{j=1}^{n} w_j = 1$. In our model, the linguistic quantifier *'usually'* is used to derive the weights of OWA aggregation. 'Usually' is a proportional linguistic quantifier Q which can be represented as a fuzzy subset over $I = [0,1]$, i.e. $Q(r)$ indicates the degree to which any proportion $r \in I$ satisfies the concept 'usually'. Moreover, 'usually' is a regular increasing monotone quantifier [19] which satisfies the following conditions:

- $Q(0) = 0$ and $Q(1) = 1$.
- If $x \geqslant y$ then $Q(x) \geqslant Q(y)$

The weights for this type of quantifiers [19] are:

$$w_j = Q\left(\frac{j}{n}\right) - Q\left(\frac{j-1}{n}\right) \tag{29}$$

This type of aggregation also protects the reputation system from being adversely affected by malicious parties who provide false or misleading information "ballot stuffing" [1].

3.2.3. A reliability measure

As we mentioned earlier the computed values of reciprocity and individual reputation scores are not always reliable and therefore a reliability measure is required to answer the question of how reliable is the computed value? An answer can be found based on Dempster–Shafer Theory (DST) which deals with the inherited ignorance/uncertainty in knowledge. Actually, DST was first introduced by Dempster in the 1960s [20] and subsequently extended by Shafer in 1976 [21]. The first step in applying DST is to define a set of possible states which is called the *frame of discernment*, denoted by Θ. The Θ delimits a set of possible states of a given system, exactly one of which is assumed to be true at any one time. Each state $B \in 2^{\Theta}$ is associated with a basic probability mass $m(B)$ such that [22]:

$$m(B) \geqslant 0 \quad \text{and} \quad m(\varnothing) = 0$$
$$\sum_{B \in 2^{\Theta}} m(B) = 1 \tag{30}$$

where 2^{Θ} is the power set of Θ. DST allows us to represent ignorance by assigning a basic probability mass to a statement which contains several alternatives or unclear outcome. By this way we can account for a lack of observation and uncertainty, both of which represent forms of incomplete information. In this direction, the Θ for our application is the set of ratings partner a_i has given to partner a_j, i.e. $\Theta = S_{a_i}(a_j) = S^{+}_{a_i}(a_j) \cup S^{-}_{a_i}(a_j) \cup S^{0}_{a_i}(a_j)$. The positive encounters $B_1 = |S^{+}_{a_i}(a_j)|$, and the negative encounters $B_2 = |S^{-}_{a_i}(a_j)|$ represent a clear outcome whereas the neutral encounters, $B_3 = |S^{0}_{a_i}(a_j)|$, represent uncertain and incomplete answer and therefore uncertainty about the judgment.

Actually, the reliability measure is inversely proportional to the uncertainty and therefore can be computed based on how much uncertainty is there in the system. Let A_j be the proposition *'positive and negative ratings partner a_j has given to partner a_i'*, then the Shafer interval will be $[s(A_j), p(A_j)]$, which is a subinterval of the unit interval. $s(A_j)$ is called the support and $p(A_j)$ is the plausibility of A_j. The support for A_j is the total mass attributed to A_j and all of its states [22].

$$s(A_j) = \sum_{B | B \subseteq A_j} m(B)$$
$$p(A_j) = \sum_{B | B \cap A_j \neq \varnothing} m(B) \tag{31}$$

The uncertainty of a proposition is the width of the Shafer interval, i.e. $u(A_j) = p(A_j) - s(A_j)$. Since $\Theta = S_{a_i}(a_j)$ and $p(A_j) = 1$ then the support is:

$$s(A_j) = m(B_1) + m(B_2) = \frac{|S^{+}_{a_i}(a_j) + S^{-}_{a_i}(a_j)|}{|S_{a_i}(a_j)|} \tag{32}$$

The reliability measure, $\text{reliab}_{a_i}(a_j)$, of a_j as seen by a_i is defined by:

$$\text{reliab}_{a_i}(a_j) = 1 - u(A_j) = 1 - m(\Theta) = 1 - \frac{|S^{0}_{a_i}(a_j)|}{|S_{a_i}(a_j)|} \tag{33}$$

Here $m(\Theta)$ represents the residual uncertainty of the domain (neutral encounters). The reliability measure for the reciprocity value needs a new *frame of discernment* to cover the uncertainty for both partners. Hence Θ for this case is $\Theta = S_{a_i}(a_j) \cup S_{a_j}(a_i)$ and therefore the reliability measure is:

$$\text{reliab}(a_i, a_j) = 1 - m(\Theta) = 1 - \frac{|S^{0}_{a_i}(a_j)| + |S^{0}_{a_j}(a_i)|}{|S_{a_i}(a_j)| + |S_{a_j}(a_i)|} \tag{34}$$

4. Trust and reputation models for movie recommender system

Recently, RS have become an essential part of e-commerce online services where they offer suggestions about products customers might also like to buy. Their contribution comes in two forms, either to predict ratings of products that a user wants to know about, or to list products that users might find of interest based on their stated preferences, online shopping choices, and the purchases of people with similar tastes or demographics. This would create new revenue opportunities and increasing both customer retention and the number of shoppers who become buyers. The most frequently used information filtering technique for RS is the *collaborative filtering* (CF) [23,24], which computes personalized recommendations by comparing multiple customers' ratings and finding those with similar tastes. Formally, CF recommender system [25] has a set of users $A = \{a_i\}_{i=1}^{M}$ rating a set of items $S = \{s_i\}_{k=1}^{K}$. Each user a_i has rated a subset of items S_i, where the rating of user a_i for item s_k is denoted by $r_{i,k}$. Depending on the user ratings, a user can be correlated with others who provide him with a set of recommendations. However, the correlation similarity alone may not be enough to guarantee high quality predictions and recommendations. Therefore, there is a need to consider other factors to enhance the prediction accuracy. Like many authors [1,3–5], we think that trust and reputation concepts are the most important factors to achieve enhanced accuracy. Our idea is to enhance the quality of neighboring users by relying only on those neighbors having high trust and reputation scores.

This section attempts to employ the proposed fuzzy trust and reputation models for movie RS. MovieLens dataset [26] consists of a set of users rating a set of movies. However, it is not clear how encounters take place and therefore some pre-processing is required to make virtual encounters. Let us assume that an encounter is formed between two partners (users) if they have rated the same movie. In the following we will explore how to implement reputation and trust models for movie RS. In our work, trust and reputation values are automatically inferred from the rating database of the RS and thereafter these values are used to enhance the accuracy of the recommendation process.

4.1. The implementation of fuzzy trust model

Some preliminary steps are required to make the MovieLens dataset suitable to implement the proposed trust and reputation models for movie RS. Actually, in real e-commerce interactions, the partner rating for a given encounter with a specific partner reflects the rater satisfaction about the encounter. Furthermore, this rating gives an image for the other counter from the point of view of the rater. Therefore, we need each partner to express his opinion for the other partner to compute reciprocity between them. To capture this image, we will use the exploratory protocol in which the consumer (active user) asks the producers (users in the historical data of RS) about the things known to him. Thus the consumer can evaluate producers based on their answer. Accordingly each consumer hides his actual ratings for the co-rated movies and asks the producer to predict them. The following Resnick's prediction formula [25] is generally used to predict a rating for a given movie

$$\text{pr}_{x,k} = m_x + \frac{\sum_{a_y \in C_N} \text{sim}(a_x, a_y) \times (r_{y,k} - m_y)}{\sum_{a_y \in C_N} |\text{sim}(a_x, a_y)|} \tag{35}$$

Here, $\text{pr}_{x,k}$ is the predicted rating user a_x may give to movie s_k, C_N is the set of neighbors, and m_y is the average of a_y's ratings that is given by:

$$m_y = \frac{1}{|S_y|} \sum_{s_k \in S_y} r_{y,k} \tag{36}$$

As mentioned earlier, reciprocity is a one-to-one relationship and therefore $|C_N| = 1$, i.e. only one producer is asked to predict the consumer ratings. Consequently, Formula (35) will be reduced to the following:

$$p_y(r_{x,k}) = (m_x - m_y) + r_{y,k} \tag{37}$$

where $r_{y,k}$ is the actual rating a_y has given to s_k and $p_y(r_{x,k})$ is the a_x's rating for s_k as predicted by a_y (the sole neighbor for a_x). At this point, we are ready to compute rating a_x has given to a_y for an encounter e_k, $r^{e_k}_{a_x}(a_y)$, based on how well a_y predicts ratings of a_x. Therefore, $r^{e_k}_{a_x}(a_y)$ is computed for $Z_B = \{-2, \ldots, +2\}$ using the following mapping:

$$r^{e_k}_{a_x}(a_y) = \begin{cases} +2 & 0.0 \leqslant |p_y(r_{x,k}) - r_{x,k}| \leqslant 0.5 \\ +1 & 0.5 < |p_y(r_{x,k}) - r_{x,k}| \leqslant 1.0 \\ 0 & 1.0 < |p_y(r_{x,k}) - r_{x,k}| \leqslant 1.5 \\ -1 & 1.5 < |p_y(r_{x,k}) - r_{x,k}| \leqslant 2.0 \\ -2 & \text{otherwise} \end{cases} \tag{38}$$

For example, if actual rating is $r_{x,k} = 3$ and the predicted rating is $p_y(r_{x,k}) = 3.3$ then Formula (38) would give $r^{e_k}_{a_x}(a_y) = +2$ and in case $r_{x,k} = 3$ and $p_y(r_{x,k}) = 1.4$ then $r^{e_k}_{a_x}(a_y) = -1$. Having computed user ratings for encounters, the trust measure can be obtained according to Formulae (6)–(23).

4.2. The implementation of fuzzy beta reputation model

We applied leave-one-out method over the users in the historical data of RS to find a reputation score for each individual in the community. The out user will be called consumer and the remaining will be called producers (the community members). Each producer hides his actual ratings and asks the consumer to predict using Formula (37). From these ratings we can form encounters according to Formula (38) and compute $N^{+d}_{x,y}$ and $N^{-d}_{x,y}$. The reputation scores from individuals are computed using Formulae (24)–(26) and the overall reputation score for a given consumer will be the aggregation of all the scores coming from the whole community. The OWA guided by a linguistic quantifier *'usually'* is used for the aggregation process. The fuzzy linguistic quantifier *usually* [27] is shown in Fig. 2 and is defined as below:

$$Q(r) = \begin{cases} 0 & 0.0 \leqslant r \leqslant 0.3 \\ 2r - 0.6 & 0.3 < r \leqslant 0.8 \\ 1 & 0.8 < r \leqslant 1.0 \end{cases} \tag{39}$$

where r is the proportion of the reputation scores that satisfies the concept 'usually'. Based on Formula (39), the weights for OWA operator can be found using Formula (29). The overall reputation score is:

$$rep(a_c) = \sum_{j=1}^{n} w_j \times rep_{\sigma(j)}(a_c) \tag{40}$$

where $\mathrm{rep}_{\sigma(1)} \leqslant \mathrm{rep}_{\sigma(2)} \leqslant \cdots \leqslant \mathrm{rep}_{\sigma(n)}$.

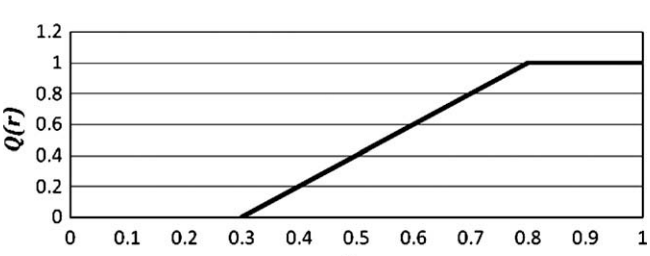

Fig. 2. Fuzzy linguistic quantifier 'usually' [27].

4.3. Implementation of eBay and beta reputation models

The proposed models are compared with the existing reputation models like eBay and beta reputation models. eBay model deploy three ratings, positive (+1), negative (−1), and neutral (0). According to Formula (38), the positive and negative rating sets, $S^+_{a_i}(a_j)$ and $S^-_{a_i}(a_j)$, for eBay model are:

$$\begin{aligned} S^+_{a_i}(a_j) &= \{r^{e_k}_{a_i}(a_j) \in \{+1, +2\} | e_k \in H_i\} \\ S^-_{a_i}(a_j) &= \{r^{e_k}_{a_i}(a_j) \in \{-1, -2\} | e_k \in H_i\} \end{aligned} \tag{41}$$

The overall numbers of positive and negative ratings according to Formula (5) are:

$$N^+_y = \sum_{a_j \in A} |S^+_{a_j}(a_y)| \quad \text{and} \quad N^-_y = \sum_{a_j \in A} |S^-_{a_j}(a_y)| \tag{42}$$

Accordingly, the eBay reputation score is $\mathrm{rep}(a_y) = N^+_y - N^-_y$. For beta reputation model the following mapping is used to meet the binary beta rating scale:

$$r^{e_k}_{a_x}(a_y) = \begin{cases} +1 & |p_y(r_{x,k}) - r_{x,k}| \leqslant 1.25. \\ -1 & \text{otherwise} \end{cases} \tag{43}$$

Accordingly, the overall numbers of positive and negative ratings (Formula (42)) are used to compute the overall beta reputation score using Formula (4).

4.4. The proposed methodology for incorporating trust and reputation in RS

Now the question arises on how should users include the trust and/or reputation concepts in the decision process of RS? Some of the previous systems incorporate either trust or reputation [4,6,13,15,28] while others treat one of them as a sub-property of the other [8]. However, by incorporating both concepts separately, the system would benefit to a considerable extent from both the concepts. For example, reputation scores of the community members will play an important role for partners who have never transacted before to filter out those having low reputation scores. On the other hand, trust measures based on the previous direct interactions (private knowledge about the trustee) can overrule the reputation score that a given partner may have. A novel methodology is presented in this work by incorporating both concepts separately following the natural way of reasoning when deciding to interact with other people. Firstly, the set of neighboring users is formed by weighting them according to their reputation (collecting neighbors based on their correlation coefficient and reputation scores). Secondly, the trust measures are used to refine the set of neighbors to get a set of recommenders who are trusted highly by the active user. This can be done by two-level filtering methodology, namely reputation-based similarity and trust-based filtering as illustrated by the block diagram in Fig. 3.

4.4.1. Reputation-based similarity

Perhaps the simplest way to incorporate reputation into the recommendation process is to combine reputation scores and correlation values to produce a compound similarity so that Resnick's formula can be used. This method is proposed by O'Donovan and Smyth [4] but under the name of trust-based weighting. The Pearson correlation coefficient [29] to compute the correlation between users, based only on the rating of the co-rated items, S_{xy}, both users have declared, is:

$$\mathrm{corr}(a_x, a_y) = \frac{\sum_{s_k \in S_{xy}} (r_{x,k} - m_x)(r_{y,k} - m_y)}{\sqrt{\sum_{s_k \in S_{xy}} (r_{x,k} - m_x)^2 \sum_{s_k \in S_{xy}} (r_{y,k} - m_y)^2}} \tag{44}$$

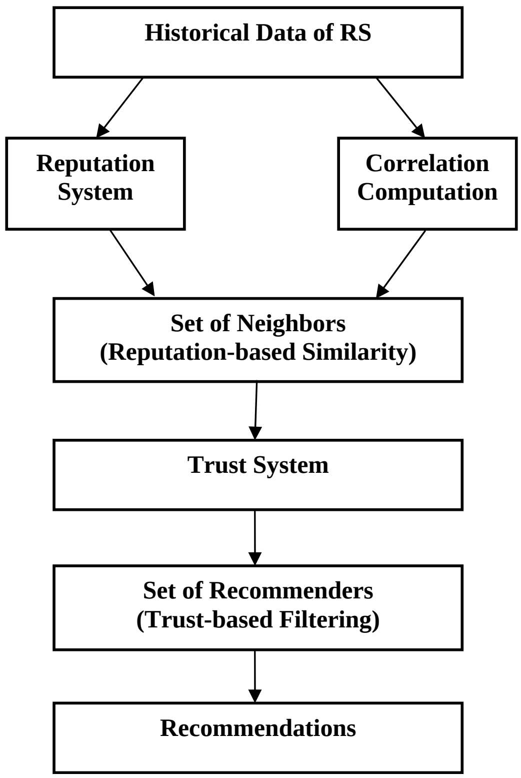

Fig. 3. The proposed methodology for incorporating trust and reputation in RS.

The following formula gives the evolved similarity measure between partners a_x (the active partner) and a_y based on the correlation value and reputation score.

$$\mathrm{sim}(a_x, a_y) = \frac{2 \times \mathrm{corr}(a_x, a_y) \times \mathrm{rep}(a_y)}{\mathrm{corr}(a_x, a_y) + \mathrm{rep}(a_y)} \tag{45}$$

4.4.2. Trust-based filtering

In this level, the trust is used as a means for filtering neighbors prior to recommendation so that only the most trustworthy neighbors participate in the recommendation process. Thus the trust system will be applied only on the set of neighbors to extract recommenders with high trust measures. This will provide a second layer of refinement for RS so that only the trustworthy neighbors could be selected as recommenders.

5. Experiments

The MovieLens dataset consists of 100,000 ratings, assigned by 943 users on 1682 movies. All ratings follow the 1-bad, 2-average, 3-good, 4-very good, and 5-excellent numerical scale. Each user has rated at least 20 movies. In our experiments, we considered only users who have rated at least 59 movies, 20 (34%) for user profiling and 39 (66%) for testing. Out of 943 users, 500 users satisfied this condition and contributed 84,773 ratings out of 100,000. This dataset is used as the basis for generating five random splits, Split-1, Split-2, Split-3, Split-4, and Split-5, where for each split, 50 users were randomly selected as active users and the remaining 450 users were utilized to generate a set of neighbors for a given active user. Thus all the experiments are repeated three times, once with each split. During the testing phase, each active user's ratings are divided randomly into two disjoint sets, training ratings (34%) and test ratings (66%). The training ratings are used as available explicit ratings from the active user, whereas the test ratings are treated as unseen ratings that the system would attempt to predict.

The effectiveness of RS is measured by the accuracy of the predictions that it makes. In this work, two evaluation metrics are used to evaluate the effectiveness of different RS, namely the mean absolute error (MAE), and the total coverage of the system. MAE measures the deviation of predictions generated by the RS from the true ratings specified by the active user. The total MAE over all the active users [29], N_T (in our experiments $N_\mathrm{T} = 50$) is calculated by:

$$\mathrm{MAE} = \frac{1}{N_\mathrm{T}} \sum_{i=1}^{N_\mathrm{T}} \left(\frac{1}{|S_i|} \sum_{k=1}^{|S_i|} |pr_{i,k} - r_{i,k}| \right) \tag{46}$$

Lower MAE corresponds to more accurate predictions of a given RS. On the other hand, coverage is the measure of the percentage of items for which a RS can provide predictions. The RS may not be able to make predictions for every item. Low coverage value indicates that the RS will not be able to assist the user with many of the items he has not rated. We compute Coverage as the number of items for which the recommender system can generate predictions over the total number of unseen items [30,31].

$$\mathrm{Coverage} = \frac{\sum_{i=1}^{N_\mathrm{T}} p_i}{\sum_{i=1}^{N_\mathrm{T}} |S_i|} \tag{47}$$

Here, p_i is the total number of predicted items for the active user a_i. Before evaluating the accuracy of our approaches, we have to build up the reputation scores for the users in the historical data of RS without reference to the set of active users. It is worth noting that ordinarily the reputation scores do not need to be computed on-the-fly because it is the community opinion about a person while trust measures would be built on-the-fly during the normal operation of RS. Having built the reputation scores the effectiveness of our approaches is evaluated by generating rating prediction for each movie in the active user test set by using the users in the historical data of RS as recommenders. In order to test the importance of incorporating trust and reputation concepts into RS based on the proposed models, five experiments are conducted according to five different scenarios. The first scenario implements a classical correlation based RS. In the second scenario, the eBay reputation model is incorporated for the RS while the beta reputation model is incorporated in the third scenario. The fuzzy beta reputation model is incorporated without and with the fuzzy trust model in the last two scenarios, respectively. The five experiments for each split are:

(1) Classical correlation-based RS (CBRS/C).
(2) Correlation-based RS with eBay reputation model (CBRS/eBay).
(3) Correlation-based RS with beta reputation model (CBRS/beta).
(4) Correlation-based RS with fuzzy reputation model (CBRS/R).
(5) Correlation-based RS with fuzzy reputation and trust models (CBRS/RT).

For CBRS/C experiment, if the correlation coefficient is negative then a_x and a_y are negatively correlated, i.e. each user discourages the other. Therefore, only positive correlation values are considered. Experiments 2, 3, and 4 use Formula (45) to compute the similarity between partners according the specified reputation model without trust-based filtering. The system picks the movies, from the test ratings set of the active user, one by one and thereafter predicts ratings for them using Formula (35) over the set of all neighbors who have rated the same movie. Having computed the predicted ratings, the system compares them with the actual

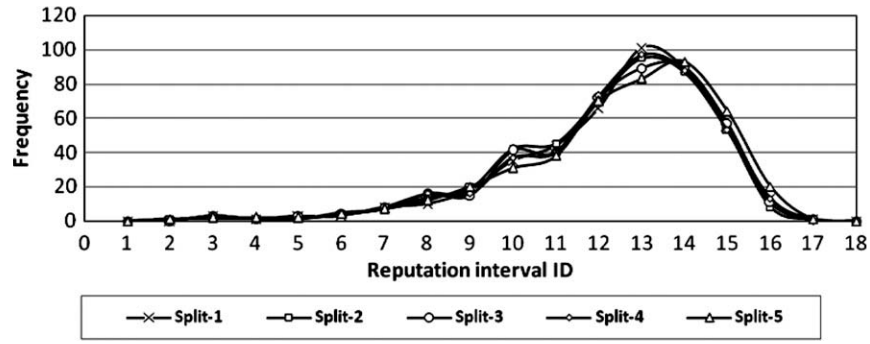

Fig. 4. Beta reputation scores distribution for the producers of all splits.

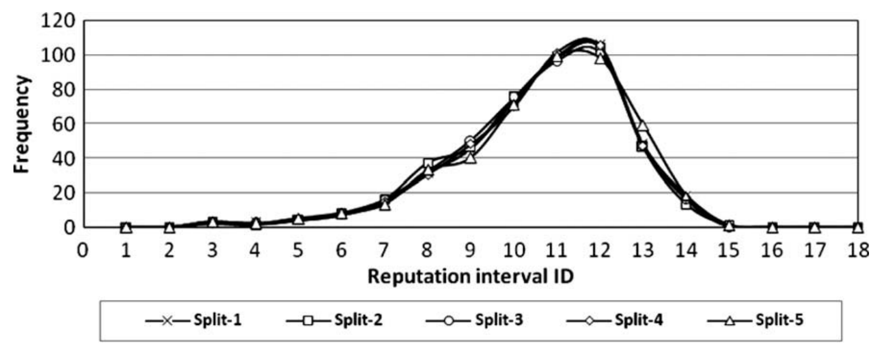

Fig. 5. Fuzzy beta reputation scores distribution for the producers of all splits.

ratings given by the active user. The set of recommenders is kept 50 for all the experiments while the neighborhood set size is kept 80.

5.1. Analysis of the results

To visualize the type of reputation scores generated by beta and fuzzy beta reputation models, we plot the distribution of beta and fuzzy beta reputation scores of the producers (users in the historical data of RS) of all splits in Figs. 4 and 5, respectively. Both fuzzy beta and beta reputation scores have a negatively-skewed distribution with a long lower tail. Taking the producers of Split-1 as an example (Table 1), most of the beta reputation scores range from 0.15 to 0.5 covering 90.67% of the population while the fuzzy beta reputation scores range from 0.1 to 0.4 covering 89.11% of the population. With fuzzy beta reputation model, approximately 3.55% of the producers have reputation scores less than 0.05% and 4.22% of the producers have reputation scores greater than 0.40. on the other hand with beta reputation model, approximately 4% of the producers have reputation scores less than 0.10% and 3.11% of the producers have reputation scores greater than 0.50.

Table 1
Reputation intervals used for reputation distribution in Figs. 4 and 5 with the corresponding data for Split-1 dataset

Interval ID	Reputation interval	Beta reputation scores		Fuzzy beta reputation scores	
		Frequency	Relative frequency	Frequency	Relative frequency
1	[−0.25, −0.20[	0	0.0000	0	0.0000
2	[−0.20, −0.15[	0	0.0000	0	0.0000
3	[−0.15, −0.10[	3	0.0067	3	0.0067
4	[−0.10, −0.05[	1	0.0022	2	0.0044
5	[−0.05, +0.00[	3	0.0067	4	0.0089
6	[+0.00, +0.05[	3	0.0067	7	0.0156
7	[+0.05, +0.10[	8	0.0178	14	0.0311
8	[+0.10, +0.15[	10	0.0222	31	0.0689
9	[+0.15, +0.20[	18	0.0400	45	0.1000
10	[+0.20, +0.25[	35	0.0778	73	0.1622
11	[+0.25, +0.30[	44	0.0978	99	0.2200
12	[+0.30, +0.35[	66	0.1467	105	0.2333
13	[+0.35, +0.40[	101	0.2244	48	0.1067
14	[+0.40, +0.45[	89	0.1978	18	0.0400
15	[+0.45, +0.50[	55	0.1222	1	0.0022
16	[+0. 50, +0.55[	13	0.0289	0	0.0000
17	[+0.55, +0.60[	1	0.0022	0	0.0000
18	[+0.60, +0.65[	0	0.0000	0	0.0000

Table 2
MAE for CBRS/C, CBRS/eBay, CBRS/beta, CBRS/R, and CBRS/RT for all splits

Split	CBRS/C	CBRS/eBay	CBRS/beta	CBRS/R	CBRS/RT
Split-1	0.761491	0.765003	0.739887	0.724541	0.712748
Split-2	0.734596	0.749935	0.725450	0.701584	0.689988
Split-3	0.761773	0.773752	0.728786	0.722702	0.712175
Split-4	0.759542	0.769807	0.751482	0.745473	0.719211
Split-5	0.763784	0.782836	0.754076	0.746971	0.730415

Table 3
Coverage for CBRS/C, CBRS/eBay, CBRS/beta, CBRS/R, and CBRS/RT for all splits

Split	CBRS/C	CBRS/eBay	CBRS/beta	CBRS/R	CBRS/RT
Split-1	94.13241	91.951361	95.580004	96.564370	97.915460
Split-2	92.79448	89.639724	95.165204	96.117230	97.591936
Split-3	93.98522	91.682420	95.772469	97.319127	98.333047
Split-4	90.35988	87.070510	91.988276	93.974923	95.782446
Split-5	93.78214	91.242013	94.966495	96.259935	97.740377

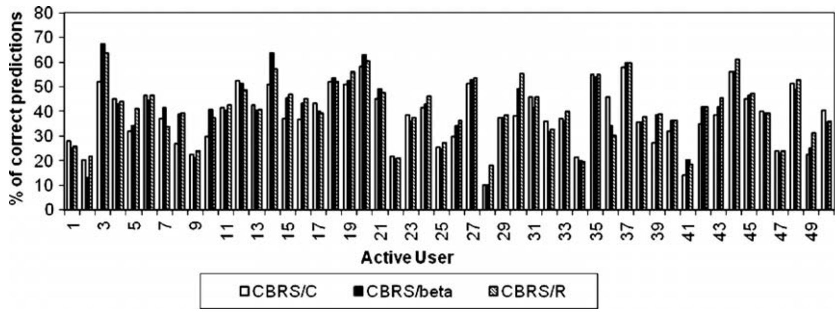

Fig. 6. Percentage of the correct predictions by CBRS/C, CBRS/beta, and CBRS/R for the active users of Split-1 dataset.

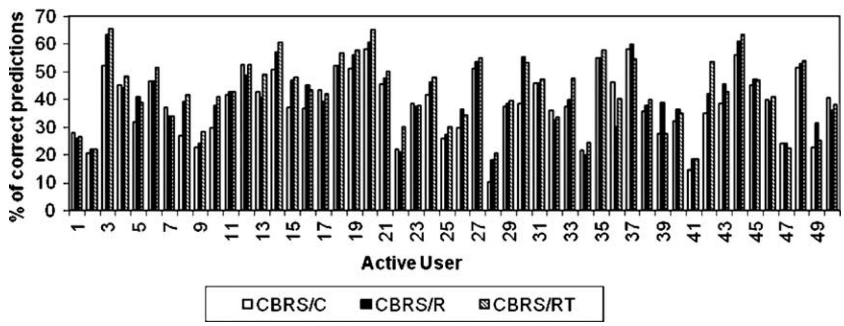

Fig. 7. Percentage of the correct predictions by CBRS/C, CBRS/R, and CBRS/RT for the active users of Split-1 dataset.

Table 4
Comparison of CBRS/R with CBRS/C, CBRS/eBay, and CBRS/beta in terms of percentage of the correct predictions for the active users of all splits

Split	CBRS/R with CBRS/C			CBRS/R with CBRS/eBay			CBRS/R with CBRS/beta		
	Greater	Same	Smaller	Greater	Same	Smaller	Greater	Same	Smaller
Split-1	32	5	13	32	4	14	29	9	12
Split-2	31	5	14	38	3	9	32	6	12
Split-3	32	7	11	37	4	9	29	4	17
Split-4	26	6	18	34	4	12	26	9	15
Split-5	26	5	19	33	3	14	31	3	16

The total MAE for CBRS/RT is smaller than that of CBRS/C and CBRS/R, while the total Coverage for CBRS/RT is greater than both CBRS/C and CBRS/R as shown in Table 2 and Table 3, respectively. Fig. 6 shows the correct prediction percentage obtained from CBRS/C, CBRS/beta, and CBRS/R for all the 50 active users of Split-1. Each graph shows the percentage of the number of ratings that the system predicted correctly out of the total number of available test ratings by the active user. The correct prediction percentage obtained from CBRS/C, CBRS/R, and CBRS/RT for the same set of active users is also given in Fig. 7. Results summarized in Table 4 show that CBRS/R outperforms CBRS/C, CBRS/eBay, and CBRS/beta for all the splits. The higher number of predictions by CBRS/R obviously illustrate that better set of like-minded users is found and therefore the accuracy gets enhanced. For example, the correct predictions generated by CBRS/R are better than or equal to that of CBRS/C on 78% cases for the best split (Split-3) whereas it is 62% for the worst split (Split-5). Also the correct predictions generated by CBRS/R are better than or equal to that of CBRS/beta on 76% cases for the best splits (Split-1 and Split-2) whereas it is 62% for the worst split (Split-3). All these improvements are achieved without increasing the cost of online processing cost which remains almost the same. CBRS/C and CBRS/R differ in the offline process complexity (the reputation scores are built offline regularly) as shown in Table 5. On the other hand, Table 6 shows that further improvements in the results are achieved by employing trust besides reputation as a second layer of filtering. For example, the predictions generated by CBRS/RT for Split-1 and Split-3 are better than or equal to 78% cases of CBRS/R. This happens because of cancelling neighbors who have low trust measures from the recommenders set. CBRS/RT outperforms both CBRS/eBay and CBRS/beta by a big margin for all the splits (Table 6). The offline process for CBRS/RT and CBRS/R regularly computes the reputation score of each individual and therefore the scores could increase or decrease according to the active user recent behavior with others.

Table 5
Important complexity parameters

RS	Offline Process Complexity	Space Complexity	Online Process Complexity
CBRS/C	–	$O(MK)$	$O(MK)$
CBRS/R	$O(M^2K)$	$O(M(K+1))$	$O(M(K+1))$
CBRS/RT	$O(M^2K)$	$O(M(K+1))$	$O(M(K+1)) + O(\|C_N\|K) \approx O(M(K+1))$

Table 6
Comparison of CBRS/RT with CBRS/R, CBRS/eBay, and CBRS/beta in terms of percentage of the correct predictions for the active users of all splits

Split	CBRS/RT with CBRS/R			CBRS/RT with CBRS/eBay			CBRS/RT with CBRS/beta		
	Greater	Same	Smaller	Greater	Same	Smaller	Greater	Same	Smaller
Split-1	35	4	11	40	4	6	37	5	8
Split-2	23	10	17	39	3	8	34	4	12
Split-3	31	8	11	41	2	7	35	4	11
Split-4	29	7	14	42	3	5	32	3	15
Split-5	31	5	14	39	2	9	33	6	11

6. Conclusion

General computational models are proposed for reputation and trust in this paper. The proposed models can be implemented for any application as long as it is a rating system explicitly or implicitly. We have used our models in building movie RS to enhance the recommendation accuracy. Incorporating trust and reputation concepts separately gives a two-level filtering methodology to enhance the recommendation accuracy through reputation-based similarity and trust-based filtering. A comparison of experimental results against those obtained using eBay and beta reputation models clearly indicates superiority of the proposed models. Our main focus in the present work was to incorporate trust and reputation concepts into movie RS and therefore sparsity and scalability problems have not been considered. However, one promising direction for future work would be to incorporate trust and reputation mechanisms together with efficient ways to overcome the sparsity and scalability problems. A compact user model [32,33] seems to be a good solution for decreasing the effect of sparsity for movie RS and thereby enhancing the system's scalability.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments.

References

[1] A. Jøsang, R. Ismail, C. Boyd, A survey of trust and reputation systems for online service provision, Decision Support Systems 43 (2) (2007) 618–644.
[2] S. Ries, J. Kangasharju, M. Muhlhauser, A classification of trust systems, in: OTM Workshops, LNCS 4277, 2006, pp. 894–903.
[3] J. Sabater, C. Sierra, Review on computational trust and reputation models, Artificial Intelligence Review 24 (2005) 33–60.
[4] J. O'Donovan, B. Smyth, Trust in recommender systems, in: Proceedings of IUI'05, San Diego, CA, USA, 2005, pp. 167–173.
[5] M. AL-Mutairi, Preference uncertainty and trust in decision making, Ph.D. Thesis, University of Waterloo, Waterloo, Ontario, Canada, 2007.
[6] J. Cho, K. Kwon, Y. Park, Collaborative filtering using dual information sources, IEEE Intelligent Systems 22 (3) (2007) 30–38. Special issue on recommender systems.
[7] S. Ruohomaa, L. Kutvonen, Trust management survey, in: P. Herrmann, et al. (Eds.), iTrust 2005, LNCS 3477, Springer-Verlag, Berlin, Heidelberg, 2005, pp. 77–92.
[8] L. Mui, M. Mohtashemi, A. Halberstadt, A computational model of trust and reputation, in: Proceedings of the 35th Hawaii International Conference on System Sciences (HICSS-35'02), 2002.
[9] D.H. McKnight, N.L. Chervany, The meaning of trust, Technical Report, Management Information Systems Research Centre, University of Minnesota, 1996.
[10] S. Marsh, Formalising trust as a computational concept, Ph.D. Thesis, Department of Mathematics and Computer Science, University of Stirling, Scotland, 1994.
[11] D. Gambetta, Can we trust trust?, in: D. Gambetta (Ed.), Trust: Making and Breaking Cooperative Relations, University of Oxford Press, 1990, pp. 213–237.
[12] E. Ostrom, A behavioral approach to the rational-choice theory of collective action, American Political Science Review 92 (1) (1998) 1–22.
[13] G. Zacharia, Collaborative reputation mechanisms for online communities, Master Thesis, Massachusetts Institute of Technology, 1999.
[14] P. Resnick, R. Zeckhauser, Trust among strangers in internet transactions: Empirical analysis of eBay's reputation system, in: M.R. Baye (Ed.), The Economics of the Internet and E-Commerce, Advance in Applied Microeconomics, vol. 11, Elsevier Science, 2002.
[15] C. Dellarocas, The digitalization of Word-Of-Mouth: promise and challenges of online reputation mechanisms, Management Science 49 (10) (2003) 1407–1424.
[16] A. Jøsang, R. Ismail, The beta reputation system, in: Proceedings of the 15th Bled Electronic Commerce Conference, Bled, Slovenia, 2002.
[17] G. Casella, R.L. Berger, Statistical Inference, Duxbury Press, 1990.
[18] L.A. Zadeh, A computational approach to fuzzy quantifiers in natural languages, Computer and Mathematics with Applications 9 (1) (1983) 149–184.
[19] R.R. Yager, On ordered weighted averaging aggregation operators in multi-criteria decision making, IEEE Transactions on Systems, Man and Cybernetics 18 (1) (1988) 183–190.
[20] A.P. Dempster, A generalization of Bayesian inference, Journal of the Royal Statistical Society, Series B 30 (1968) 205–247.
[21] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
[22] P. Lucas, L. Van Der Gaag, Principles of Expert Systems, Addison-Wesely Publishing Company, 1991.
[23] D. Goldberg, D. Nichols, B.M. Oki, D. Terry, Using collaborative filtering to weave an information tapestry, Communication of the ACM 35 (12) (1992) 61–70.
[24] B. Miller, I. Albert, S. Lam, J. Konstan, J. Riedl, MovieLens unplugged: experiences with an occasionally connected recommender system, in: Proceedings of the International Conference on Intelligent User Interfaces, 2003.
[25] G. Adomavicius, A. Tuzhilin, Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions, IEEE Transaction on Knowledge and Data Engineering 17 (6) (2005) 734–749.
[26] http://www.movielens.umn.edu.
[27] L.A. Zadeh, A note in web intelligence, world knowledge and fuzzy logic, Data and Knowledge Engineering 50 (2004) 149–184.
[28] P. Massa, P. Avesani, Trust-aware collaborative filtering for recommender systems, in: Proceedings of the Move to Meaningful Internet Systems (CoopIS), DOA, and ODBASE, LNCS 3290, Springer, 2004, pp. 492–508.
[29] J.S. Breese, D. Heckerman, C. Kadie, Empirical analysis of predictive algorithms for collaborative filtering, in: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, Madison, WI. Morgan Kaufmann, San Francisco, CA, 1998, pp. 43–52.
[30] E. Vozalis, K.G. Margaritis, Analysis of recommender systems' algorithms, in: Proceedings of the sixth Hellenic-European Conference on Computer Mathematics and its Applications (HERCMA), Athens, Greece, 2003.
[31] J.L. Herlocker, L.A. Konstan, L.G. Terveen, J.T. Riedl, Evaluating collaborative filtering recommender systems, ACM Transaction on Information Systems 22 (1) (2004) 5–53.
[32] Mohammad Yahya H. Al-Shamri, K.K. Bharadwaj, Fuzzy-genetic approach to recommender systems based on a novel hybrid user model, Expert Systems with Applications 35 (2008) 1386–1399.
[33] Mohammad Yahya H. Al-Shamri, K.K. Bharadwaj, A compact user model for hybrid movie recommender system, in: Proceedings of ICCIMA'07, IEEE Computer Society, India, 2007, pp. 519–524.