Ratings in Distributed Systems: A Bayesian Approach

Lik Mui, Mojdeh Mohtashemi, Cheewee Ang, Peter Szolovits
MIT Laboratory for Computer Science
200 Technology Square, Cambridge, MA 02139
{lmui, mojdeh, cheewang, psz}@lcs.mit.edu

Ari Halberstadt
Magiccookie
9 Whittemore Road, Newton, MA 02458
ari@magiccookie.com

Abstract

For distributed systems at large and e-commerce systems in particular, ratings play an increasingly important role. Ratings confer reputation measures about sources. This paper reports our formalization of the rating process. This paper argues that ratings should be context- and individual- dependent quantities. In contrast to existing rating systems in many e-commerce or developer sites, our approach makes use of personalized and contextualized ratings for assessing source reputation. Our approach is based on a Bayesian probabilistic framework.

1. Introduction

In a networked environment where resources are distributed widely, locating relevant resources has been the subject of the information retrieval industry. However, reliability measures for the retrieved data are often missing.[1] This paper reports our progress on the development of a mathematical scheme and computational tools for how approval ratings can be propagated and assimilated for the purpose of assigning reputation values to other agents. Degrees of approval are modeled in a probabilistic framework. We adopt the following definition for reputation:

- ***Reputation***: perception that an agent has of another's intentions and norms.

Discussion on the meaning of this definition can be found in [Mui, *et al.*, 2002], along with how reputation relates to concepts such as trust:

- ***Trust***: a subjective expectation an agent has about another's future behavior based on the history of their encounters.

We argue that the failure of existing rating systems is partially due to the lack of personalization and contextualization in these systems. For example, consider an expert in French cuisine browsing the Amazon.com catalogue, she comes across a French cookbook with a 3 stars rating (as provided by other raters). The average Amazon user is not likely to be a French cuisine expert. For our expert, the three star rating is therefore of limited value. Unless the rating system can summarize the ratings of those who are similar to the individual seeking a rating, the rating system is destined to provide noisy ratings. This example points to the importance of personalizing ratings for those seeking source sanctioning. Similarly, a user's reputation as a computer scientist is likely to be different from his or her reputation as a violinist. Existing systems such as those in eBay and Slashdot do not take into account this context dependency. Section 2 covers the relevant literature. Section 3 formalizes intuitions concerning ratings. Section 4 concludes this paper with thoughts on future works.

2. Background

The information retrieval (IR) community has been tackling the question of information relevance for decades now, and with remarkable success (Salton, 1971; Kleinberg, 1997; Brin and Page, 1998). However, simply locating relevant resources in distributed systems today is no longer adequate. Users demand not only relevant but also reliable information.

Collaborative filtering is a paradigm for using the collective experiences of a cyber community to select relevant resources (Goldberg, *et al.*, 1992). Large numbers of users are often required before

[1] Although different in other contexts, reliability here is taken to confer source reputation. The intuition used here is that the more reputed is a source, the more reliable is information gathered from it.

reasonable performance is achieved. No metric exists to measure how reliable is a rating nor how reputed is a rater. The recent human-computer interface (HCI)'s interest in "social navigation" highlights the importance of leveraging collective experience in aiding users interacting with information systems (Dieberger, *et al*., 2000; Wexelblat, 2001). However, techniques for assessing the reliability of the collected data are *ad-hoc*.

Prompted by the rise of on-line communities and their proliferation in the past decade, rating systems have been built to provide source reputation or reliability assessment. On-line auction eBay provides a +1/-1 rating system for buyers to provide feedback about sellers and vice versa. A few recent high profile fraud cases in eBay by individuals with high eBay ratings suggest that the company should seriously consider enhancing their simple rating system. Rating systems in such commercial sites aggregate all users' ratings equally and ignore the personalized nature of reputation (*i.e.*, one's reputation means different things to different individuals depending on the context). We term this type of reputation measure "global" in nature – as opposed to being personalized for the inquirer's context.

In the computer science literature, Zacharia and Maes (1999) have suggested that reputation in an on-line community can be related to the ratings that an agent receives from others, and have pointed out several criteria for such rating systems. However, their mathematical formulation for the calculation of reputation can at best be described as intuitive – without justifications except the intuitive appeal of the resulting reputation dynamics. *Ad-hoc* formulation plagues several other proposals for rating or reputation systems such as those in Glass, *et al*. (2000), Yu and Singh (2000), Rouchier, et al. (2001), Sabater, *et al.* (2001), Esfandiari, *et al.* (2001), among others. Nevertheless, ratings and reputation measures have been found to provide useful intuition or services for users of these systems.

3. Conceptual Framework

On notations, consider the universe of users and objects in the system as the following sets:

Set of agents: $A = \{ a_1, a_2, \ldots a_M \}$ **Set of objects**: $O = \{ o_1, o_2, \ldots o_N \}$

Users rate the objects, which could be other users (*i.e.,* $A \subset O$). The set of ratings ρ can be viewed as a mapping from A x O to a real number between 0 and 1 (to accommodate probability).

$$\textbf{Rating}: \quad \rho : A \times O \rightarrow [0, 1]$$

To model the process of opinion sharing among agents, the concept of an encounter is required. An encounter is an event between two different agents (a_i, a_j) such that the query agent a_i asks the responding agent a_j for a_j's rating of an object:

$$\textbf{Encounter}: \quad d \in D = A^2 \times O \times [0, 1] \cup \{ \perp \}$$

where $\{\perp\}$ represents the set of no encounter ("bottom").

Consider an agent a_i who has never interacted with object o_k in the past. Before taking the trouble of interacting with o_k, a_i asks other agents (A_{-i}) in the environment their ratings for o_k. a_i will decide to interact with object o_k, for example, if the weighted sum of the ratings (a proxy for o_k's reputation) from agents in A_{-i} is greater than a certain threshold. The **reputation** of a_j in a_i's mind can be considered as the probability that in the next encounter, a_i will approve of a_j's rating for the object context. A context describes a set of attributes about an environment. Let an attribute be defined as the presence ('1') or absence ('0') of a trait. The set of all attributes is possibly countably infinite, and is defined as follows:

Attribute: $b \in B$ *and* $b : O \rightarrow \{ 0, 1 \}$ **Set of attributes**: $B = \{ b_1, b_2, \ldots \}$

A context is then an element of the power set of B:

$$\textbf{Context}: \quad c \in C \text{ and } C = P\{B\}$$

where $P\{.\}$ represents the power set. The reputation mapping can now be represented by:

Reputation: $R : A \times A \times C \rightarrow [0, 1]$

This mapping assumes that reputation only changes after an encounter – either directly with the individual involved, or through rating feedbacks from other agents. After each encounter, the update rule for the state of the system can be represented as below (which is discussed in the next section):

Update rule: *newstate*: $R \times D \rightarrow R$

3.1 Bayesian Reputation Inference

Let $x_{ab}(i)$ be the indicator variable for agent *a*'s approval of *b* after the i^{th} encounter between them – thus the ratings ρ is restricted to binary values of 0 or 1. If *a* and *b* have had *n* encounters in the past for context *c*, the proportion estimator $\hat{\theta}_c$ for the number of approvals of *b* by *a* can be modeled with a Beta prior distribution (Dudewicz and Mishra, 1988):

$$p(\hat{\theta}_c) = Beta(c_1, c_2)$$

where c_1 and c_2 are parameters determined by prior assumptions and $\hat{\theta}_c$ is the estimator for the true proportion of *a*'s approval of *b* based on encounters between them for context *c*. (From this point on, a given context *c* is assumed and the subscript is avoided for clarity.) Assuming that each encounter's approval probability is independent of other encounters between *a* and *b*, the likelihood of having *p* approvals and $(n - p)$ disapprovals can be modeled as:

$$L(D \mid \hat{\theta}) = \hat{\theta}^p (1-\hat{\theta})^{n-p} \qquad where\ D = \{\ x_{ab}(1), x_{ab}(2), \ldots x_{ab}(n)\ \}$$

Combining the prior and likelihood, the posterior estimate can be shown to be (Mui, *et al.*, 2001):

$$p(\hat{\theta} \mid D) = Beta(c_1 + p, c_2 + n - p)$$

In our framework, reputation for *b* in *a*'s mind is *a*'s estimate of the probability that *a* will approve of *b* in the next encounter $x(n + 1)$. This estimate is based on *n* previous encounters between them (for the specific context *c*). This estimate can be shown to yield the following (*ibid.*):

$$p\left(x_{ab}(n+1) = 1 \mid D\right) = E\left[\hat{\theta}_n \mid D\right] = r_{ab} \tag{1}$$

This conditional expectation is the expression for r_{ab} : the update rule specified in the last section. First order statistical properties of the posterior are summarized below for the posterior estimate of $\hat{\theta}$:

$$E\left[\hat{\theta} \mid D\right] = \frac{c_1 + p}{c_1 + c_2 + n} \qquad \sigma_{\hat{\theta}|D}{}^2 = \frac{(c_1 + p)(c_2 + n - p)}{(c_1 + c_2 + n - 1)(c_1 + c_2 + n)^2}$$

which indicates an easy implementation for the estimation procedure. Note that r_{ab} in (1) is a personalized estimate by agent *a*, which depends on the set of encounters *D* that agent *a* takes into account about *b*. Specifying the set of encounters is individual dependent and could include inferred encounters from hearsays from agent *a*'s neighbors, as suggested in the propagation procedures below. We also have not specified how prior belief about $\hat{\theta}$ is formed? Two situations are considered below.

3.2 Complete Strangers: Prior Assumptions

If agents a_i and a_j are complete strangers, an ignorance assumption is made. When these two strangers first meet, their estimate for each other's reputation should be uniformly distributed across the reputation's domain:

$$p(\hat{\theta}) = \begin{cases} 1 & 0 < \hat{\theta} < 1 \\ 0 & otherwise \end{cases}$$

For the Beta prior, values of c_1=1 and c_2=1 yields such a uniform distribution.

3.3 Known Strangers: Prior Assumptions

If a_i and a_k have never met before but a_i knows a_j well (*i.e.,* a_i has an opinion on a_j's reputation). Also, a_j knows a_k well. a_i would like to estimate a_k's reputation based on a_i 's history of encounters with a_j, and a_j's history of encounters with a_k. This setup is depicted below:

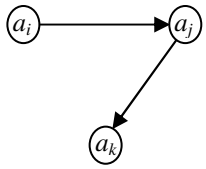

Figure 1. *Illustration of an indirect neighbor pair (a_i, a_k)*

For agent a_i, agent a_k is a stranger but a_k is not completely unknown to a_i since a_i knows a_j who has an opinion about k. In a future encounter between a_i and a_k, the probability that this encounter will be rated good by a_i is:

Let $D_{ij,n}$ = all n encounters between a_i and a_j.

$D_{jk,m}$ = all m encounters between a_j and a_k.

$x_{ij}(n)$ = indicator variable for a_i's approval of a_k at encounter n

$$\begin{aligned} & p\left(x_{ik}(n+1)=1 \mid D_{ij}, D_{jk}\right) \\ &= p\left(x_{ij}(n+1)=1 \mid D_{ij}\right) \cdot p\left(x_{jk}(n+1)=1 \mid D_{jk}\right) + \\ & \quad \left[1-p\left(x_{ij}(n+1)=1 \mid D_{ij}\right)\right] \cdot \left[1-p\left(x_{jk}(n+1)=1 \mid D_{jk}\right)\right] \\ &= r_{ij} r_{jk} + (1-r_{ij}) \cdot (1-r_{jk}) \end{aligned} \tag{2}$$

The interpretation of this equation is that the probability that a_i would approve a_k at encounter $n + 1$ is the sum of the probabilities that both a_i and a_j agree (about the context that c is in) and that both of them disagree.

3.4 Inference Propagation

So far, we have discussed belief estimation for indirect neighbors which are one degree away from a direct neighbor. The reputation of indirect neighbors which are further away can be estimated by recursively applying Equation (2) using closer neighbors as direct neighbors. This procedure works if there exists a single path between two agents. This recursion has a fixed point for a finite population community: when all members of the communities have been rated, the recursion stops.

If there are multiple paths between two agents, as illustrated in *Figure 2*, a different procedure is required.

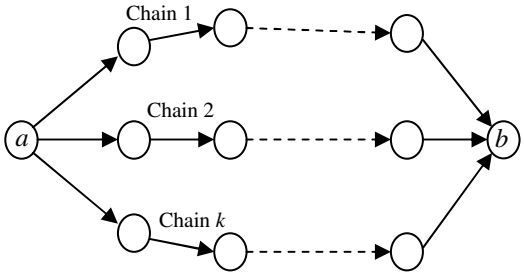

Figure 2*. Illustration of a parallel network between two agents* ***a*** *and* ***b.***

Figure 2 shows a parallel network of k chains between two agents of interest, where each chain consists of at least one link. Agent a would like to estimate agent b's reputation as defined by the embedded network between them. Clearly, to sensibly combine the evidence about b through all intermediate agents in between, measures of "reliability" about these intermediates are required. One way these measures can be calculated is shown in Mui, *et al.* (2002) by using Chernoff bound on the parameter estimate $\hat{\theta}$. Let m represents the minimum number of encounters necessary to achieve a given level of confidence δ for $\hat{\theta}$ and error bound ε; the following inequality can be demonstrated:

$$m \geq -\frac{1}{2\varepsilon^2} \ln\left((1-\delta)/2\right)$$

As δ approaches 1, a larger m is required to achieve a given level of error bound ε. This threshold (m) can be set on the number of encounters between agents such that a reliable measure can be established as follows:

$$w_{ab} = \begin{cases} \dfrac{m_{ab}}{m} & \text{if } m_{ab} < m \\ 1 & \text{otherwise} \end{cases}$$

where m_{ab} is the number of encounters between agents a and b. Intuition of this formula is as follows: m represents the minimum sample size of encounters to reach a confidence (and error) level about the parameter estimates. Above a given level of sample size, the estimator is guaranteed to yield the specified level of confidence. Therefore, such an estimate can be considered as "reliable" with respect to the confidence specification. Any sample size less than the threshold m is bound to yield less reliable estimates. As a first order approximation, a linear drop-off in reliability is assumed here.

For each chain in the parallel network, how should the total weight be tallied? Two possible methods are plausible: additive and multiplicative. The problem with additive weight is that if the chain is "broken" by a highly unreliable link, the effect of that unreliability is local to the immediate agents around it. In a long social chain however, an unreliability chain is certain to cast serious doubt on the reliability of any estimate taken from the chain as a whole. On the other hand, a multiplicative weighting has "long-distance" effect in that an unreliable link affects any estimate based on a path crossing that link. The form of a multiplicative estimate for chain i's weight (w_i) can be:

$$w_i = \prod_{j=1}^{l_i} w_{ij} \qquad \text{where } 0 \leq i \leq k$$

where l_i refers to the total number of edges in chain i and w_{ij} refers to the j^{th} segment of the i^{th} chain.

Once the weights of all chains of the parallel network between the two end nodes are calculated, the estimate across the whole parallel network can be sensibly expressed as a weighted sum across all the chains:

$$r_{ab} = \sum_{i=1}^{k} r_{ab}(i)\overline{w_i}$$

where $r_{ab}(i)$ is a's estimate of b's reputation using path i and $\overline{w_i}$ is the normalized weight of path i ($\overline{w_i}$ sum over all i yields 1). r_{ab} can be interpreted as the overall perception that a garnered about b using all paths connecting the two.

4. Discussion and Conclusion

This paper has briefly reported our Bayesian formalization of the rating process. We have argued that since ratings and reputations are clearly context dependent quantities, models about them should explicitly take the context into account. Applications of this formalization have been applied to simulated and real world ratings and are reported in Mui, *et al.*, 2001. Our Bayesian approach achieves more

accurate reliability and reputation estimates than control schemes – similar to those used in existing systems such as eBay and Amazon.

There are several yet to be resolved issues. We have examined binary ratings in this paper. Extending the Bayesian framework to other types of ratings should be straightforward. Specifying the procedure for inferring ratings and reputations from one context to another is also needed. We have started investigating the use of ontologies to relate different contexts. A hard problem is to determine how to resolve the different ontological views of the world held by different agents. Furthermore, the metric or function to transfer rating or reputation from one context to the next is yet to be worked out. In our formulation, the calculation of an agent's reputation requires the disclosure of detailed personal rating and object descriptions to other agents. This creates a privacy concern which needs to be addressed.

5. Acknowledgements

We thank the support of the MEDG group in MIT/LCS for providing a supportive research environment. We especially thank Dr. Jon Doyle for his discussions.

6. References

S. Brin, L. Page (1998) "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Proc. 1998 WWW Conference*.

A. Dieberger, P. Dourish, K. Hook, P. Resnick, and A. Wexelblat (2000) "Social Navigation: Techniques for Building More Usable Systems," *Interaction, 7*(6), pp. 36-45.

E. J. Dudewicz, S. N. Mishra (1998) *Modern Mathematical Statistics,* John Wiley & Sons.

B. Esfandiari, S. Chandrasekharan (2001) "On How Agents Make Friends: Mechanisms for Trust Acquisition," *4th Workshop on Deception, Fraud and Trust in Agent Societies*, Montreal, Canada.

A. Glass, B. Grosz (2000) "Socially Conscious Decision-Making," *Autonomous Agents'2000*.

D. Goldberg, D. Nichols, B. M. Oki, D. Terry (1992) "Using collaborative filtering to weave an information tapestry." *Communication of the ACM, 35*(12), pp. 61-70.

J. Kleinberg (1997) "Authoritative sources in a hyperlinked environment." *IBM Technical Report RJ 10076*.

L. Mui, M. Mohtashemi, C. Ang, P. Szolovits, A. Halberstadt (2001) "Bayesian Ratings in Distributed Systems: Theories, Models, and Simulations," *MIT LCS Memorandum*.

L. Mui, M. Mohtashemi, A. Halberstadt (2002) "A Computational Model of Trust and Reputation," *Proc. 34th Hawaii International Conference on System Sciences*.

J. Rouchier, M. O'Connor, F. Bousquet (2001) "The Creation of a Reputation in an Artificial Society Organized by a Gift System," *J. Artificial Societies and Social Simulations, 4*(2).

J. Sabater, C. Sierra (2001) "REGRET: A reputation Model for Gregarious Societies," *4th Workshop on Deception, Fraud and Trust in Agent Societies*, Montreal, Canada.

G. Salton (1971) *SMART Retrieval System: Experiments in Automatic Document Processing,* Englewood Cliffs, NJ: Prentice-Hall.

S. Wasserman, K. Faust (1994) *Social Network Analysis: Methods and Applications,* Cambridge University Press.

A. Wexelblat (2001) "History-Rich Tools for Social Navigation," MIT Media Lab Report.

B. Yu, M. P. Singh (2000) "A Social Mechanism of Reputation Management in Electronic Communities," *Cooperative Information Agents'2000*.

G. Zacharia, P. Maes (1999) "Collaborative Reputation Mechanisms in Electronic Marketplaces," *Proc. 32nd Hawaii International Conference on System Sciences*.