Computational Intelligence, Volume 23, Number 3, 2007

A COMPREHENSIVE APPROACH FOR SHARING SEMANTIC WEB TRUST RATINGS

JIE ZHANG AND ROBIN COHEN

David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada

In the context of the Semantic Web, it may be beneficial for a user (consumer) to receive ratings from other users (advisors) regarding the reliability of an information source (provider). We offer a method for building more effective social networks of trust by critiquing the ratings provided by the advisors. Our approach models the consumer's private reputations of advisors based on ratings for providers whom the consumer has had experience with. It models public reputations of the advisors according to all ratings from these advisors for providers, including those who are unknown to the consumer. We then combine private and public reputations by assigning weights for each of them. Experimental results demonstrate that our approach is robust even when there are large numbers of advisors providing large numbers of unfair ratings. We show that we can effectively model the trustworthiness of advisors even when the population of providers grows increasingly large and discuss how our approach is beneficial in modeling providers. As such, we present a framework for sharing ratings of possibly unreliable sources, of value as users on the Semantic Web attempt to critique the trustworthiness of the information they seek.

Key words: trust on the Semantic Web, content of Web sources, information sharing, unfair ratings, Web of Trust.

1. INTRODUCTION

The vision of the Semantic Web is to construct a common semantic interpretation for World Wide Web pages, to one day reliably run software to interpret the information conveyed in any of its documents. In building the Semantic Web, however, information may be supplied by a wide selection of sources, with the result that a user seeking information will need to judge whether the content of any given source is in fact trustworthy. It is, therefore, important to develop models for trust in the context of the Semantic Web. Various approaches to date have been formulated about how best to form a Web of Trust, to share information and selectively choose trustworthy partners from whom information may be obtained. In our research, we are considering a problem that arises when social networks are formed to share trust ratings—that of unfair ratings. Dellarocas (2000) distinguishes unfair ratings as unfairly high ratings and unfairly low ratings. Unfairly high ratings may be used to increase the trustworthiness of others and promote their services. They are often referred to as "ballot stuffing." Unfairly low ratings of others are often referred to as "bad-mouthing." In brief, the ratings of the trustworthiness of others, obtained from third parties, may in fact be suspect. What is required therefore is a mechanism for effectively adjusting the basis on which decisions of trust are made, to discount these possibly unfair ratings.

In this paper, we discuss our research in the context of sharing ratings of sources (called information providers) among users on the Semantic Web. We present an approach for modeling the trustworthiness of advisors—those users providing trust ratings for potential providers from whom information may be obtained. We refer to the user seeking advice as the consumer. We first represent private reputation values, based on what is known about the advisors' ratings for providers with whom the consumer has already had some experience. We then describe how to construct a public model of trustworthiness of advisors based on common, centrally held knowledge of providers and the ratings provided by advisors, including the trust ratings of providers totally unknown to the consumer. We then outline how both private and public models can be combined, to obtain a value for the trustworthiness of each possible advisor. In summary, we offer a method for building more effective social networks of trust, by critiquing the advice provided by advisors.

In Section 2, we introduce the Semantic Web setting for sharing information about sources, and present some current research on modeling the trustworthiness of information sources based on ratings provided by advisors. Section 3 presents our approach for modeling the trustworthiness of advisors according to the ratings provided by them in the context of the Semantic Web. Section 4 provides examples that go through each step of our approach and carefully draw attention to some of the valuable features of our model. Section 5 includes some experimental results demonstrating what happens when there are large numbers of advisors providing large numbers of unfair ratings and showing the ability of our approach to operate effectively in environments with growing numbers of providers. We also present results to demonstrate the effectiveness of our approach for modeling trustworthiness of advisors when consumers attempt to model the trustworthiness of providers, based on the ratings supplied by advisors. Conclusions and future work are outlined in Section 6.

2. BACKGROUND AND RELATED WORK

In this section, we discuss the setting of sharing information about sources, on the Semantic Web. We motivate the need to acquire information about the reliability of sources and then briefly outline some current research on modeling the trustworthiness of sources. This includes some discussion of approaches to communicate with other users to obtain advice about sources, sometimes referred to as a Web of Trust (Gil and Ratnakar 2002), as well as an approach for addressing the problem that some users may provide untruthful advice.

The challenge of trusting information providers in a Web-based environment is discussed in (Paolucci et al. 2003). Paolucci et al. provided valuable insights into the need for trust on the Web, in the context of Web services, where Web sites dynamically exchange information using XML descriptions, but where it is difficult to ensure that the meaning of the messages being sent is well understood, without human intervention. The Semantic Web contributes by providing ontologies for Web services to interpret meanings in exchanged messages. According to (Paolucci et al. 2003), with the Semantic Web, the interaction between users and providers needs a process of capability matching to link users with providers of Web services. Specifically, providers advertise their capabilities, a user sends a request for the type of service he requires, a registry matches the capabilities of providers and the capabilities expected by the user, and finally the user selects the most suitable provider. However, in their advertisements, providers may lie about their capabilities in order to be selected by the user. To avoid selection of an untruthful provider, there is a need to properly model the trustworthiness of providers. In (Gil and Ratnakar 2002) this problem is reinforced for the Semantic Web: whether to trust the content of a Web resource, depending on the source. Richardson et al. (2003) explain further that due to the great diversity of the Web, it is difficult to expect the content to be consistent and of high quality. It then becomes important to decide how trustworthy each information source is.

Maximilien and Singh (2004, 2005) adopt an agent-based approach for modeling trust on the Semantic Web. Their work focuses on representing multiple qualities of services (QoS) for automatic runtime Web service selection. This trust model is based on a shared conceptualization of QoS and takes into account providers' quality advertisement, consumers' quality preferences, quality relationships, and consumers' quality tradeoffs. To select a Web service implementation, a consumer dynamically associates a trust value with each service implementation and selects the service implementation with the highest assigned level of trust. The trust value of each service implementation partially depends on its reputation value, which is determined by the set of quality values from other users who previously selected that provider.

Kagal et al. (2002) use a DAML + OIL trust ontology in a multi-agent system, which is based on a distributed trust and delegation mechanism verifying that a user's credentials are acceptable. The trust ontology is built for specifying credentials and checking if the credentials conform to policies. A policy maps credentials to a certain ability or right. The mechanism allows propagation of trust beliefs exchanged between users and avoids repeated checking of users' credentials.

The research of Gil and Ratnakar (2002) provides a framework for users to express their trust about a source and the statements the source contains, by annotating each part of the source to indicate their views. The focus of the work is on how to provide an effective interface for users to record their annotations. This TRELLIS system ultimately averages the ratings provided over many users and many analyses, to present a reflection of the trustworthiness of the source. A credibility-reliability pair emerges for each source-statement pair, to derive an overall rating of a single source, based on each of the associated statements provided by the source.

Modeling trust on the Semantic Web, as discussed so far in this section, includes a reliance on the beliefs or ratings provided by third parties to be truthful. In fact, it is important to address the problem of possibly unfair or unreliable ratings. One approach that explores this possibility is that of Richardson et al. (2003). In this work, each user first explicitly specifies a small set of users whom he trusts, leading to a Web of Trust. This arrangement allows any user to compute the trustworthiness of a possible provider, based on the ratings supplied by others in his social network. The trust value of a provider is computed locally by combining the trust ratings provided by other users. One feature of this approach is to recursively propagate trust through the user's social network. In effect, trust in a provider is derived using some aggregating functions along each possible chain of trust from the user to the provider. One concern with this approach, however, is that this method of propagating trust may be computationally intractable, as there may be many different paths, of various lengths, which need to be aggregated.

In our own research, we are developing a model for representing the reliability of advisors from whom advice may be sought, when a user seeks to evaluate the trustworthiness of a provider. This framework is sufficiently general to operate in a variety of environments including electronic commerce, where buyers may make decisions about sellers by soliciting input on those sellers from other buyers in the marketplace.

In the context of the Semantic Web, our model is useful for the problem of determining the reliability of a provider being evaluated by a consumer by virtue of trust ratings provided by advisors. Our focus is on addressing the problem of advisors who may be untrustworthy. The existence of malicious advisors is in fact acknowledged in (Richardson et al. 2003). However in contrast to the model of Richardson et al. (2003), we provide a more direct evaluation of each possible advisor in a Web of Trust, leading to an evaluation about how best to make use of that advisor's ratings of a possible provider being examined by a consumer.

As will be seen in the sections that follow, we make various limiting assumptions (which are revisited as future work) in order to examine more clearly the need to adjust for possibly unfair ratings from advisors. In particular, we do not envisage entire chains of trust from advisor to advisor, instead evaluating independently the trustworthiness of each advisor, based in part on the user's own past experience. In addition, we represent the input from each advisor as a summary rating of a possible source as simply reliable or unreliable. We also allow an advisor to rate a source several times. In so doing, we allow for dynamically varying the trustworthiness of the source. In addition, we introduce a forgetting factor which can be used to facilitate the comparison of advisor and consumer ratings for a provider, when the data is sparse. We also discuss the value of our approach in a context where consumers rely on advice from advisors when evaluating the trustworthiness of a provider.

3. MODELING TRUSTWORTHINESS OF ADVISORS

In the discussion below, we use the following terminology:

- **User/Consumer:** Person seeking information from various sources.
- **Provider:** An information source, providing information.
- **Advisors:** Other users providing ratings of providers to consumers.
- **Private reputation:** A determination of the reputation of an advisor by a user, based on commonly rated providers.
- **Public reputation:** A determination of the reputation of an advisor by a user, based on a centrally held model of the advisor, from interactions with a whole set of providers.

Our method for determining the trustworthiness of advisors is to employ a combination of what we refer to as private and public reputation values. To explain, the private reputation of an advisor is calculated by a consumer,[1] based on ratings the advisor supplies of providers with whom the consumer has already had some experience. If the advisor is reputable and has similar preferences as the consumer, the consumer and advisor will likely have many ratings in common. This can then be used as the basis for assessing the trustworthiness of the advisor. In cases where the consumer has little private knowledge of the advisor, a public reputation will be elicited, reflecting the trustworthiness of that advisor, based on her ratings of all providers in the system. A weighted combination of private and public reputations is derived, based on the estimated reliability of the private reputation value. This combined value then represents the trustworthiness of the advisor. Providers are to be rated only after an advisor has had personal experience with that provider.[2]

3.1. Private Reputation

Our approach allows a consumer C to evaluate the private reputation of an advisor A by comparing their ratings for commonly rated providers $\{P_1, P_2, \ldots, P_m\}$. For one of the commonly rated providers P_i ($1 \leq i \leq m$ and $m \geq 1$), A has the rating vector R_{A,P_i} and C has the rating vector R_{C,P_i}. A rating for P_i from C and A is binary ("1" or "0," for example), in which "1" means that P_i is trustworthy and "0" means that P_i is untrustworthy. For the purpose of simplicity, we assume ratings for providers are binary. Possible ways of extending our approach to accept ratings other than binary ones will be investigated as future work. Further discussion can be found in Section 6.

The ratings in R_{A,P_i} and R_{C,P_i} are ordered according to the time when they are provided. The ratings are then partitioned into different elemental time windows. The length of an elemental time window may be fixed (e.g., three days) or adapted by the frequency of the ratings to the provider P_i, similar to the way proposed in (Dellarocas 2000). It should also be considerably small thus that there is no need to worry about the changes of providers' behavior within each elemental time window. We define a pair of ratings (r_{A,P_i}, r_{C,P_i}), such that r_{A,P_i} is one of the ratings of R_{A,P_i}, r_{C,P_i} is one of the ratings of R_{C,P_i}, and r_{A,P_i} corresponds to r_{C,P_i}. The two ratings, r_{A,P_i} and r_{C,P_i}, are correspondent only if the rating r_{C,P_i} is the most recent rating in its time window, and the rating r_{A,P_i} is the closest and prior to the rating r_{C,P_i}. We consider ratings provided by C after those by A, to incorporate into C's ratings anything learned from A, before taking an action. According to the solution proposed by Zacharia et al.

[1] It is expected that the human user will have an agent acting on his behalf to perform these calculations.

[2] This may be kept in check by a centralized system where all consumers agree to have their interactions with providers known, for instance.

(1999), by keeping only the most recent ratings, we can avoid the issue of advisors "flooding" the system. No matter how many ratings are provided by one advisor in a time window, we only keep the most recent one.

We define the rating pair (r_{A,P_i}, r_{C,P_i}) as a positive rating pair if r_{A,P_i} is the same value as r_{C,P_i}. Otherwise, the pair is a negative rating pair. We assume that r_{C,P_i} is provided within the time window T_C and r_{A,P_i} is within the time window T_A. We assume that each time window is identified by an integer value, where 1 is the most recent time window with a rating, 2 is the time window just prior, and so on until the oldest time window. Thus, T_A is always greater than or equal to T_C because r_{A,P_i} is prior to the rating r_{C,P_i}. As also pointed out by Jøsang and Ismail (2002), old ratings may not always be relevant for providers' actual trustworthiness because providers may change their behavior over time. Older ratings should be given less weight than more recent ones. In our case, if r_{A,P_i} and r_{C,P_i} are within the same time window, it is more relevant to compare them and the rating pair will be given more weight; otherwise, the rating pair will be given less weight. We calculate the weight of the rating pair, (r_{A,P_i}, r_{C,P_i}), as follows:

$$z = \lambda^{T_A - T_C}, \tag{1}$$

where λ is a forgetting factor (a concept used in (Jøsang and Ismail 2002)) and $0 \leq \lambda \leq 1$. Note that when $\lambda = 1$ there is no forgetting (i.e., older ratings supplied by advisors will be accepted and compared to the consumer's rating of the closest time window), and when $\lambda = 0$ only the rating pair with ratings that are within the same time window will be considered. In cases where C and A always provide ratings within the same time window, the value of $T_A - T_C$ is always 0, thus that the weight of the rating pair is always 1. Note as well that when $\lambda > 0$, the higher the value of λ, the greater the weight placed on the ratings provided by the advisor.

We examine rating pairs for all commonly rated providers. We define N_p as the sum of the weights of all positive rating pairs and N_n as the sum of the weights of all negative rating pairs for all commonly rated providers. The private reputation of the advisor A is estimated as the probability that A will provide reliable ratings to C. Because there is only incomplete information about the advisor, the best way of estimating the probability is to use the expected value of the probability. The expected value of a continuous random variable is dependent on a probability density function, which is used to model the probability that a variable will have a certain value. Because of its flexibility and the fact that it is the conjugate prior for distributions of binary events (Russell and Norvig 2002), the beta family of probability density functions is commonly used to represent probability distributions of binary events (see, e.g., the generalized trust models BRS (Jøsang and Ismail 2002) and TRAVOS (Teacy et al. 2005)). Therefore, the private reputation of A can be calculated as follows:

$$\alpha = N_p + 1, \beta = N_n + 1$$

$$R_{pri}(A) = E(\Pr(A)) = \frac{\alpha}{\alpha + \beta}, \tag{2}$$

where $\Pr(A)$ is the probability that A will provide fair ratings to C,[3] and $E(\Pr(A))$ is the expected value of the probability.

[3]An advisor's rating is considered to be a fair rating if it is the same as the consumer's rating. The consumer may decide not to trust the advisor if they have a different view of providers.

3.2. Public Reputation

When there are not enough rating pairs, the consumer C will also consider A's public reputation. The public reputation of A is estimated based on her ratings and other ratings for the providers rated by A. Each time A provides a rating $r_{A,P}$, the rating will be judged centrally as a consistent or inconsistent rating. We define a rating for a provider as a consistent rating if it is consistent with the majority of the ratings of the provider up to the moment when the rating is provided.[4] We consider only the ratings within a time window prior to the moment when the rating $r_{A,P}$ is provided, and we only consider the most recent rating from each advisor. In so doing, as providers change their behavior and become more or less trustworthy to each advisor, the majority of ratings will be able to change.

Suppose that the advisor A totally provides N_{all} ratings. If there are N_c number of consistent ratings, the number of inconsistent ratings provided by A will be $N_{\text{all}} - N_c$. In a similar way as estimating the private reputation, the public reputation of the advisor A is estimated as the probability that A will provide consistent ratings. It can be calculated as follows:

$$\alpha' = N_c + 1, \beta' = N_{\text{all}} - N_c + 1$$

$$R_{\text{pub}}(A) = \frac{\alpha'}{\alpha' + \beta'}, \tag{3}$$

which also indicates that the more the percentage of consistent ratings advisor A provides, the more reputable she will be considered.

3.3. Trustworthiness

To estimate the trustworthiness of advisor A, we combine the private reputation and public reputation values together. The private reputation and public reputation values are assigned different weights. The weights are determined by the reliability of the estimated private reputation value.

We first determine the minimum number of rating pairs needed for C to be confident about the private reputation value he has of A. The Chernoff Bound theorem (Mui et al. 2002) provides a bound for the probability that the estimation error of private reputation exceeds a threshold, given the number of pairs. Accordingly, the minimum number of pairs can be determined by an acceptable level of error and a confidence measurement as follows:

$$N_{\text{min}} = -\frac{1}{2\varepsilon^2}\ln\frac{1-\gamma}{2}, \tag{4}$$

where ε is the maximal level of error that can be accepted by C, and γ is the confidence measure. If the total weight of all rating pairs is larger than or equal to N_{min}, consumer C will be confident about the private reputation value estimated based on his ratings and the advisor A's ratings for all commonly rated providers. Otherwise, there are not enough rating pairs, the consumer will not be confident about the private reputation value, and he will then also consider public reputation. The reliability of the private reputation value can be measured as follows:

$$w = \begin{cases} \dfrac{N_p + N_n}{N_{\text{min}}} & \text{if } N_p + N_n < N_{\text{min}}; \\ 1 & \text{otherwise.} \end{cases} \tag{5}$$

[4]Determining consistency with the majority of ratings can be achieved in a variety of ways, for instance averaging all the ratings and seeing if that is close to the advisor's rating, which is the method used in our experiments in Section 5.

TABLE 1. Ratings of Providers Provided by Advisors

A_j	A_x					A_y					A_z				
T	T_1	T_2	T_3	T_4	T_5	T_1	T_2	T_3	T_4	T_5	T_1	T_2	T_3	T_4	T_5
P_1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
P_2	1	1	1	1	1	0	1	0	1	1	0	0	0	0	0
P_3	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0
P_4	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0
P_5	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0

The trust value of A will be calculated by combining the weighted private reputation and public reputation values as follows:

$$Tr(A) = wR_{\text{pri}}(A) + (1 - w)R_{\text{pub}}(A). \tag{6}$$

It is obvious that the consumer will consider less the public reputation value when the private reputation value is more reliable. Note that when $w = 1$, the consumer relies only on private reputation.[5]

4. EXAMPLES

To illustrate how our approach models trustworthiness of advisors, this section provides examples that go through each step of the approach. Examples are also provided to demonstrate how trust values different consumers have of the same advisors may vary, and to show the effectiveness of our approach even when the majority of ratings are unfair. We provide a further example to show that the forgetting factor in our model is beneficial when ratings provided by consumers and advisors are sparse.

In the setting of sharing information on the Semantic Web, a provider P_0, who is an information source, provides some information. Whether a consumer C can trust this information depends on how much C trusts P_0. To model the trustworthiness of the provider P_0, the consumer C seeks advice from three advisors A_x, A_y, and A_z who have had experience with P_0. The advice about P_0 from A_x, A_y, and A_z are ratings representing the trustworthiness of P_0. Before aggregating the ratings provided by A_x, A_y and A_z, the consumer C needs to evaluate the reliability of those ratings, which depends on the trustworthiness of the advisors A_x, A_y, and A_z. Our approach effectively models the trustworthiness of advisors based on how reliable the previous ratings provided by them are.

Consider the case where the advisors A_x, A_y, and A_z each has rated only the five providers (P_1, P_2, P_3, P_4, and P_5). Table 1 lists the ratings provided by A_j ($j \in \{x, y, z\}$) for the five providers. The symbol "T" represents a sequence of time windows, in which T_1 is the most recent time window. To simplify the demonstration, we assume that each advisor provides at most one rating within each time window. We also assume that those are the only ratings provided by them.

[5] This can be used as well if the majority rating is suspect. The consumer can rely on his own private knowledge and allow for a difference of opinion. Once a consumer has had personal experience, he will know better whether the majority opinion is acceptable.

TABLE 2. Ratings Provided by the Consumer C

T	T_1	T_2	T_3	T_4	T_5
P_1	1	1	1	1	1
P_2	1	1	1	1	–
P_3	1	1	1	–	–
P_4	1	1	–	–	–
P_5	1	–	–	–	–

TABLE 3. Private and Public Reputation Values of Advisors

A_j	A_x	A_y	A_z
$N_p(A_j)$	15	8	0
$N_n(A_j)$	0	7	15
α	16	9	1
β	1	8	16
$R_{pri}(A_j)$	0.94	0.53	0.06
$N_c(A_j)$	25	12	0
α'	26	13	1
β'	1	14	26
$R_{\text{pub}}(A_j)$	0.96	0.48	0.04

As can be seen from Table 2, the consumer C has also provided some ratings for the five providers. The consumer C might have not provided any rating for some providers within some time window. For example, C has provided only one rating for the provider P_5, which is in the time window T_1. We assume that the ratings provided by C are after those provided by A_x, A_y, and A_z if they are within the same time window.

We compare the ratings provided by A_x, A_y, and A_z in Table 1 and ratings provided by C in Table 2. The consumer C has different numbers of $N_p(A_j)$ positive and $N_n(A_j)$ negative rating pairs with A_x, A_y, and A_z, which are listed in Table 3. Accordingly, as can be seen from Table 3, the private reputation values of A_x, A_y, and A_z are different, in which the private reputation value of A_x is the highest and that of A_z is the lowest. Note that the private reputation values of advisors are calculated by setting λ to be 0, meaning that we compare only the ratings provided by C and advisors that are within the same time windows. The result indicates that the advisor A_x is most likely to provide fair ratings and have similar preferences with the consumer C, whereas A_z most likely will lie and have different preferences with C.

According to Table 1, the total number of ratings provided by each advisor is the same ($N_{\text{all}}(A_j) = 25$). We also count the number of consistent ratings each advisor provides, $N_c(A_j)$. A rating here is considered as a consistent rating when it is consistent with the majority of ratings for the provider within a same time window. Consider the case where all of the five providers are trustworthy and the majority of ratings are fair. In this situation, ratings consistent with the majority are fair. A rating of 1 provided by an advisor will be considered as a rating consistent with the majority rating, whereas a rating of 0 will be considered as an inconsistent rating. From the advisors' ratings listed in Table 1, we can see

TABLE 4. Trustworthiness of Advisors

ε	0.1	0.15	0.2
$N_{\min}$	115	51	29
w	0.13	0.29	0.52
$\text{Tr}(A_x)$	0.957	0.954	0.950
$\text{Tr}(A_y)$	0.487	0.495	0.506
$\text{Tr}(A_z)$	0.043	0.046	0.05

that ratings provided by the advisor A_x are all consistent with the majority rating, the advisor A_z always provides inconsistent ratings, and some of the ratings provided by the advisor A_y are consistent. Table 3 lists the number of consistent ratings provided by each advisor and the corresponding public reputation value of her. From Table 3, it is clear that the advisor A_x is most likely to provide consistent and therefore fair ratings, and the advisor A_z most likely will provide inconsistent ratings.

To combine private reputation and public reputation, the weight w should be determined. The value of w depends on the values of ε and γ, and the total number of rating pairs, which can be calculated as $N_p(A_j) + N_n(A_j)$ and is the same for every advisor in our example. Suppose we have a fixed value, 0.8 for γ, which means that the confidence value should be no less than 0.8 in order for the consumer to be confident with the private reputation values of advisors. In this case, the more errors he can accept, the more confident he is with the private reputation values of advisors, which also means that the more weight he will put on the private reputation values. Table 4 lists different acceptable levels of errors, their correspondent weights of private reputation values, and different results of trust values. It clearly indicates that A_x is the most trustworthy, and A_y is more trustworthy than A_z. As a result, the consumer C will place more trust in the advice provided by A_x. C will consider the advice provided by A_x more heavily when aggregating the advice provided by A_x, A_y, and A_z for modeling the trustworthiness of the provider P_0. Our framework serves the purpose of representing the trustworthiness of advisors, thus that this may be taken into account, when determining how heavily to rely on their advice.

To demonstrate how the trust values different consumers have of the same advisors may vary, we consider another consumer C', who also needs to make a decision on whether to trust the information provided by a provider P'_0 (P'_0 may differ from P_0). The ratings provided by C' for the five providers are listed in Table 5. By going through the same process as above, we can calculate the trust values the consumer C' has of A_x, A_y, and A_z, when $\varepsilon = 0.2$ and $\gamma = 0.8$. The results are presented in Table 6. Comparing Table 6 with Tables 3 and 4, we can see that the private reputations the consumer C' has of advisors are different from those the consumer C has. Although the public reputations of advisors that the consumers have are the same, the trust values that the consumers have are still different.

To show the robustness of our model, we now consider a case where the majority of ratings provided by advisors are unfair. Adjusting our earlier example, a rating of 1 provided by an advisor for any provider will now be considered as an inconsistent rating with low reputability, whereas a rating of 0 will be considered as a consistent rating. As a result, the public reputations that the consumer C has of the advisors A_x, A_y, and A_z will be different, which can be seen from Table 7. We model the trust values the consumer C has of the advisors A_x, A_y, and A_z, when C's acceptable levels of errors of private reputation values are different. Results are presented in Table 8. From this table, we can see that our approach can

TABLE 5. Ratings Provided by the Consumer C'

T	T_1	T_2	T_3	T_4	T_5
P_1	1	1	–	–	1
P_2	1	–	–	1	–
P_3	1	1	–	–	–
P_4	1	1	–	–	–
P_5	1	–	–	–	–

TABLE 6. Trust Values C' Has of Advisors

A_j	A_x	A_y	A_z
$R_{pri}(A_j)$	0.92	0.58	0.08
$R_{\text{pub}}(A_j)$	0.96	0.48	0.04
$\text{Tr}(A_j)$	0.947	0.514	0.054

TABLE 7. Public Reputations of Advisors When Majority of Ratings Are Unfair

A_j	A_x	A_y	A_z
$N_c(A_j)$	0	13	25
α'	1	14	26
β'	26	13	1
$R_{\text{pub}}(A_j)$	0.04	0.52	0.96

TABLE 8. Trustworthiness of Advisors When Majority of Ratings Are Unfair

ε	0.1	0.2	0.25
N_{min}	115	29	19
w	0.13	0.52	0.79
$\text{Tr}(A_x)$	0.157	0.508	0.751
$\text{Tr}(A_y)$	0.521	0.525	0.528
$\text{Tr}(A_z)$	0.843	0.492	0.249

still correctly represent the trustworthiness of advisors by making adjustments to rely more heavily on the private reputations.

We set the forgetting factor λ to be 0 in the above examples, meaning that we compare only the ratings provided by consumers and advisors that are within the same time windows. However, when ratings provided by them are sparse, consumers may set λ to be other values, to gain more private knowledge about advisors and rely on it more heavily when modeling trustworthiness of advisors. We use a simple example here to demonstrate how the forgetting factor in our approach is beneficial for consumers. In this example, a consumer C and an

TABLE 9. Ratings of P_1' and P_2' Provided by C and A

	P_1'						P_2'					
T	T_1	T_2	T_3	T_4	T_5	T_6	T_1	T_2	T_3	T_4	T_5	T_6
A	–	1	–	1	–	1	–	1	–	1	–	1
C	1	–	1	–	1	–	1	–	1	–	1	–

TABLE 10. Private Reputation of A and Its Weights for Different λ Values

λ	0	0.5	1
$N_p + N_n$	0	3	6
$R_{pri}(A)$	0.5	0.8	0.875
w	0	0.16	0.32

advisor A both have provided some ratings for the information providers P_1' and P_2', as listed in Table 9. We can see that the consumer C and the advisor A do not have ratings in the same time windows.

In this example, when modeling the trustworthiness of advisor A, we have $N_{\min}$ of 19, by setting ε to be 0.25 and γ to be 0.8. We also assume that each subsequent time window is one unit apart from the previous one, thus that $T_A - T_C = 1$. By setting different values for λ, we then calculate the corresponding private reputation of the advisor and the value w in the calculation of the trustworthiness of the advisor that represents how much the consumer will rely on the private reputation. These values are listed in Table 10. From this table, we can see that there are no ratings to be compared with if we set λ to be 0. By setting λ to be higher, the consumer can have more sense about the advisor, and therefore rely more on his private knowledge of the advisor.

5. EXPERIMENTAL RESULTS

Our approach models the trustworthiness of advisors according to how reliable the ratings provided by them are. To demonstrate the effectiveness of the approach, we carry out some preliminary experiments involving advisors who provide different percentages of unfair ratings. The expectation is that trustworthy advisors will be less likely to provide unfair ratings, and vice versa. We also examine how large numbers of dishonest advisors (i.e., advisors who provide unfair ratings) will affect the estimation of advisors' trustworthiness. Results indicate that our approach is still effective by making adjustments to rely more heavily on private reputations of advisors, in this case. We conduct further experiments to test the scalability of our approach. Results show that trustworthiness of advisors remains nearly the same for different populations of providers. We also demonstrate how consumers can effectively model trustworthiness of providers, making use of advisors' models created through our approach.

The first experiment involves 100 providers, 3 consumers, and 1 advisor. The 3 consumers, C_1, C_2, and C_3, rate 10, 40, and 70 randomly selected providers, respectively. The advisor totally rates 40 randomly selected providers.[6] We examine how the trust values the consumers

[6]Note that we simplify the experiments by limiting each consumer or advisor to provide at most one rating for each provider.

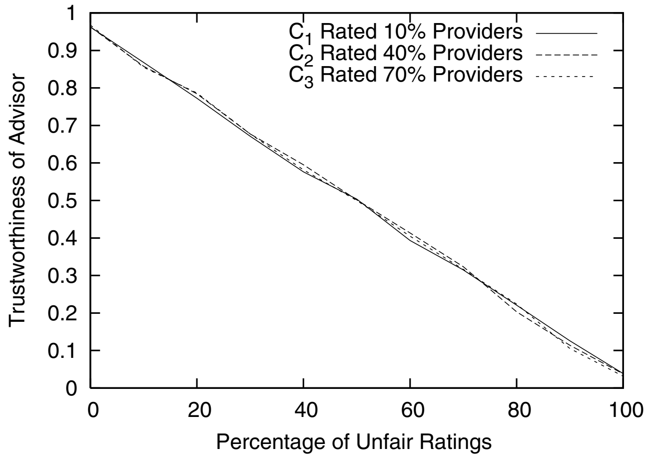

FIGURE 1. Trustworthiness of advisor.

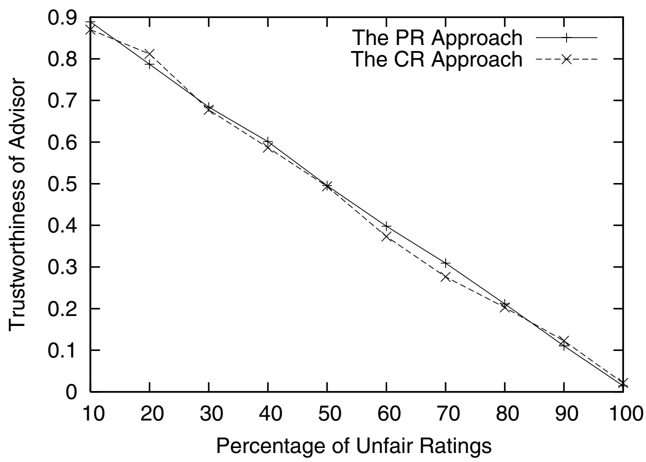

FIGURE 2. Trustworthiness of A when majority of advisors are honest.

have of the advisor change when different percentages (from 0% to 100%) of the advisor's ratings are unfair. As illustrated in Figure 1, the trust values the consumers have of the advisor decrease when more percentages of the advisor's ratings are unfair. From this figure, we can also see that our approach is still effective when the consumer C_1 does not have much experience with providers, in the sense that C_1 can still reduce the trustworthiness of the advisor when the advisor provides more unfair ratings.

The second experiment involves 100 providers, 80 advisors, and 1 consumer. The consumer and each advisor rate 80 of the randomly selected providers. We model the trust value the consumer has of one of the advisors, A. The trustworthiness of the advisor will be modeled as the combination of her private and public reputations (referred to as the CR approach) and as only her public reputation (referred to as the PR approach), respectively. The advisor A will provide different percentages (from 10% to 100%) of unfair ratings. Figure 2 illustrates the trustworthiness of A when 24 (30% of all) advisors are dishonest. Those dishonest advisors provide the same percentage of unfair ratings as the advisor A does. Results indicate

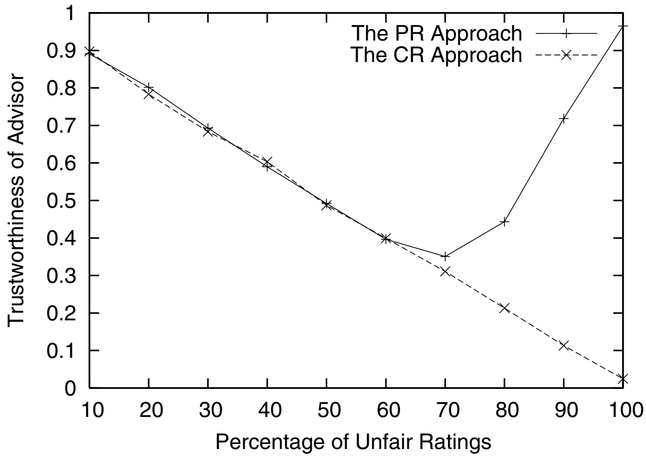

FIGURE 3. Comparison of the CR and PR approaches.

that the trustworthiness of A modeled by using the CR and PR approaches decreases when more percentages of ratings provided by A are unfair. Therefore, these two approaches are not affected when only a small number of advisors are dishonest. Figure 3 represents the trustworthiness of A when 48 (60% of all) advisors are dishonest. In this figure, the trustworthiness of A modeled by using the CR approach still decreases when more percentages of ratings provided by A are unfair, which indicates that our approach is still effective when the majority of advisors provide large numbers of unfair ratings. In contrast, the trustworthiness modeled by using the PR approach increases when more than 60% of ratings provided by the dishonest advisors are unfair, which indicates that the PR approach is only effective when the majority of ratings are fair.

The effectiveness of our approach is demonstrated by the above experiments with the fixed population of (100) providers. It is useful to examine whether our approach will still be useful when there are a large number of providers. The number of providers affects the number of commonly rated providers, and may then affect the calculation of private reputation for advisors. More specifically, in the environment where there are many providers, there may be a smaller percentage of those providers that have been commonly rated by consumers and advisors. In this case, consumers may have less private knowledge about advisors. We use a simulation to demonstrate that our approach can still effectively model trustworthiness of advisors. In this simulation, we have different populations of providers spanning from 100 to 500 in increments of 50. A consumer models trustworthiness of an advisor. Fifty percent of the ratings provided by the advisor are unfair in this experiment. The results are shown in Figure 4. The x-axis represents the populations of providers, and the y-axis represents the trustworthiness of the advisor. The solid line is the average trust value of the advisor. As can be seen from Figure 4, the trustworthiness of the advisor remains nearly the same when the population of providers changes, which indicates that our approach is scalable.

After demonstrating the effectiveness of our approach in modeling trustworthiness of advisors, we carry out a further experiment to examine how consumers can make use of our method for modeling advisors to effectively model the trustworthiness of providers. This experiment also involves 100 providers, 80 advisors, and 1 consumer. Similarly, the consumer and each advisor rate 80 of the randomly selected providers. Every 10% of the providers acts dishonestly with different probabilities (from 0 to 0.9). The consumer models

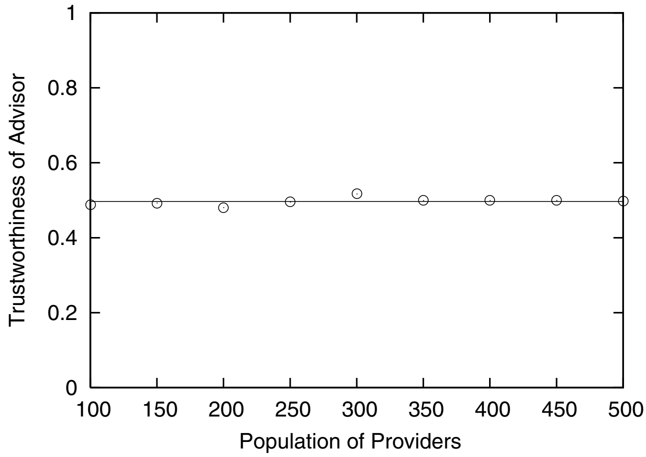

FIGURE 4. Scalability of our approach.

the trustworthiness of providers based on the advisors' ratings of providers. To determine which advisors the consumer should ask advice from, the consumer first models trustworthiness of advisors, and then selects a list of trustworthy advisors from whom he can ask advice about providers. Once this list is determined, the ratings of each of the advisors in the list need to be combined to determine the trustworthiness of the providers. For this experiment, we assume that the 10 most trustworthy advisors are kept in the list. We also adopt the aggregation function proposed by Jøsang and Ismail (2002), which combines ratings through the beta family of probability density functions, discounted by the trustworthiness of the advisors. The method also weights more heavily more recent ratings of providers and as such fits well with our particular approach for modeling trustworthiness.[7]

Similar to the second experiment, the trustworthiness of each advisor will be modeled based on either the CR approach or the PR approach. Figure 5 illustrates the trustworthiness of different providers when 30% of advisors are dishonest. Results indicate that the trustworthiness of providers, when using the CR and PR approaches to model trustworthiness of advisors, decreases when they act dishonestly with higher probabilities. Therefore, these two approaches are both effective when only a small number of advisors are dishonest. Figure 6 represents the trustworthiness of providers when 60% of advisors are dishonest. In this figure, the trustworthiness of providers, when using the CR approach to model trustworthiness of advisors, still decreases when the providers act dishonestly in higher probabilities, which indicates that our approach is still effective when the majority of advisors provide large numbers of unfair ratings. In contrast, the trustworthiness of providers, when using the PR approach to model trustworthiness of advisors, increases when the providers act dishonestly in higher probabilities. This indicates that the PR approach is only effective when the majority of ratings are fair. All in all, if taking our model and using it as a basis for evaluating providers, more accurate decisions about trustworthiness of providers can be made than using other methods for modeling advisors.

[7]Note that other methods may be used to determine the list of trustworthy advisors to consult (for example, using a threshold and retaining only advisors with trustworthiness beyond that threshold). In addition, other aggregation functions could be introduced as well. Our model is able to operate effectively with many different methods for these design decisions.

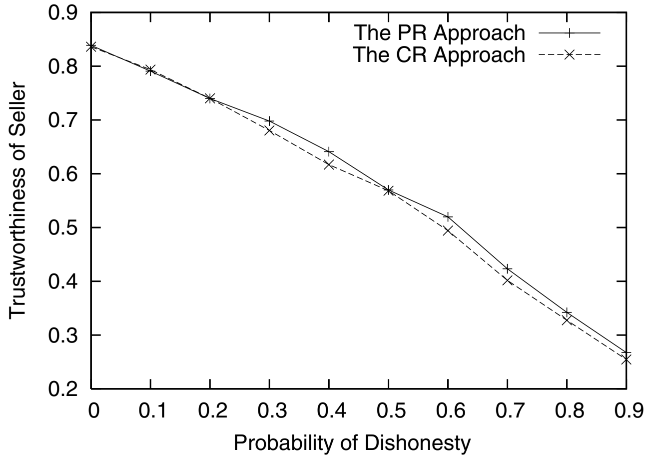

FIGURE 5. Trustworthiness of providers when majority of advisors are honest.

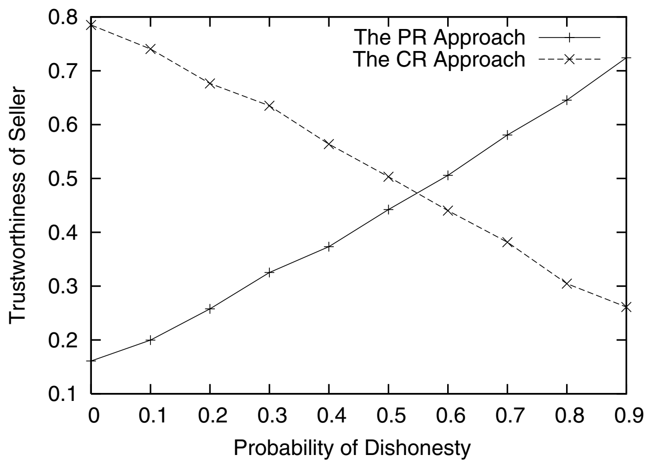

FIGURE 6. Comparison of the CR and PR approaches.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we first introduce the Semantic Web setting for sharing information about sources. Due to the fact that any user on the Web can become an information source, there is a need to form a Web of Trust. Some current research on modeling the trustworthiness of information sources on the Semantic Web relies on the unrealistic assumption that advice provided by advisors about an information source is truthful. A useful method to address this problem is to critique advisors' advice based on their trustworthiness. We present an approach for modeling the trustworthiness of advisors. Our approach allows a consumer to estimate the trustworthiness of an advisor based on the advisor's ratings for providers with whom the consumer has already had some experience. It also models the trustworthiness of the advisor based on all of her ratings and on common knowledge of providers who might be totally unknown to the consumer. We then propose combining these results to determine the overall trustworthiness of an advisor. We validate our approach by carrying out experiments in the setting where advisors may provide different numbers of unfair ratings. Experimental results indicate that our approach can effectively model the trustworthiness of advisors even when consumers do not have much experience with providers. Also, our approach is still effective when the majority of advisors provide large numbers of unfair ratings. Furthermore, our approach is scalable in terms of different populations of involved providers. We also demonstrate how our approach is helpful when used by consumers to evaluate the trustworthiness of a provider.

Our approach of combining both private and public reputation values offers useful improvement for the modeling of the trustworthiness of advisors. Other research has been conducted on this topic within the multi-agent systems community. Sabater and Sierra (2005) offers an overview of some of the earliest trust and reputation modeling systems. For instance, the REGRET system (Sabater and Sierra 2002) proposes that the trustworthiness of advisors be determined by a combination of individual, social and ontological trust measures. There are also other systems that are closer to our own research, specifically modeling the trustworthiness of advisors to determine whether to make use of that advice, using probabilistic reasoning. A model such as BRS (Jøsang and Ismail 2002) that relies on public reputation has the problem that it is only effective when the majority of ratings are fair, whereas a model like TRAVOS (Teacy et al. 2005) that uses private reputation has difficulty when a consumer is new to the system.

For the purpose of simplicity, the current approach limits ratings for providers to be binary. In future work, we will extend our approach to accept ratings in different ranges. Instead of using the numerical difference of two ratings, comparison of the two ratings could take into account the semantics of rating levels (Chen and Singh 2001). For example, although the numerical differences of the pairs are same, the difference between “5” (very trustworthy) and “3” (neutral) is smaller than that between “4” (trustworthy) and “2” (untrustworthy). In consequence, the similarity between “5” and “3,” say 0.2, should be set to be larger than the similarity between “4” and “2,” say 0. When these extensions are made, the Dirichlet family of probability density functions (Gelman et al. 2004), which is the multivariate generalization of the beta family, can be used to represent probability distributions of discrete similarity values. Our model will evaluate private and public reputation values based on aggregation of those discrete similarity values.

Our approach represents trustworthiness of providers using a single rating provided by consumers or advisors. For future work, as in the research of (Richardson et al. 2003), we will also extend our approach to accept multiple ratings representing different dimensions of trustworthiness of providers. We could for example, examine credibility and reliability of providers as used by Gil and Ratnakar (2002) or a quality of service ontology used by Maximilien and Singh (2004, 2005). We would then need to explore methods to combine the different kinds of ratings provided by advisors, for example whether to weight one dimension more heavily than another.

Another valuable direction for future work is to go beyond a generalized trust rating for an information source, to one that determines whether to trust a source on a particular topic or segment of information provided by the source. In this case, we would want to model the advisors’ trustworthiness with respect to these segments of the source, as well. This may result in the design of a more elaborate private reputation model or a method of determining what weight to place on this private reputation, when advisors have only currently rated different segments of the source. It would also be valuable to learn which advisors to rely on, for which different elements of a source.

It would also be interesting to examine how our approach can be robust when advisors choose to strategically provide truthful ratings for some providers and untruthful ratings for other providers. We first note that our approach is effective in relying on the public reputation of advisors when the majority of advisors are trustworthy. When the majority are in fact untrustworthy, however, our approach can still effectively model the trustworthiness of advisors by relying more on the private reputation of advisors, as consumers gain more experience with their advice. The experimental results in this paper have shown some aspects of this argument (see Figures 3 and 6). For future work, we plan to carry out more extensive experiments to determine how well our approach can cope with advisors who are colluding with specific providers in supplying untruthful ratings.

It is important to note that we are focused in this paper on the question of judging the trustworthiness of advisors, as part of the process of evaluating how much to trust the content of an information source. We have some initial findings in evaluating the usefulness of the model to evaluate providers. In fact, we would like to see our approach integrated into a full scale decision-theoretic framework for selecting trustworthy sources. The performance of the overall system would then need to be evaluated, along a number of different dimensions as well. We will also carry out further experiments to continue to compare our model with competing approaches. It would worthwhile, for example, to run direct comparisons with the BRS (Jøsang and Ismail 2002) and TRAVOS (Teacy et al. 2005)) models, to determine whether the trustworthiness of the provider is determined more effectively using our model. It might also be possible to have the competing approaches operating in a real-world context, to observe the performance with respect to actual information sources.

REFERENCES

CHEN, M., and J. P. SINGH. 2001. Computing and Using reputations for internet ratings. *In* Proceedings of the Third ACM Conference on Electronic Commerce, pp. 154–162, Tampa, FL.

DELLAROCAS, C. 2000. Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior. *In* Proceedings of the Second ACM Conference on Electronic Commerce, pp. 150–157, Minneapolis, MN.

GELMAN, A., J. B. CARLIN, H. S. STERN, and D. B. RUBIN. 2004. Bayesian Data Analysis. Chapman & Hall/CRC, Boca Raton, FL.

GIL, Y., and V. RATNAKAR. 2002. Trusting information sources one citizen at a time. *In* Proceedings of the First International Semantic Web Conference, pp. 162–176, Sardinia, Italy.

JØSANG, A., and R. ISMAIL. 2002. The beta reputation system. *In* Proceedings of the Fifteenth Bled Electronic Commerce Conference, pp. 324–337, Bled, Slovenia.

KAGAL, L., T. FININ, and A. JOSHI. 2002. Developing secure agent systems using delegation based trust management. *In* Proceedings of the First International Autonomous Agents and Multi Agent Systems, pp. 27–34, Bologna, Italy.

MAXIMILIEN, E. M., and M. P. SINGH. 2004. Toward autonomic web services trust and selection. *In* Proceedings of the Second International Conference on Service Oriented Computing, pp. 212–221, New York, NY.

MAXIMILIEN, E. M., and M. P. SINGH. 2005. Agent-based trust model using multiple qualities. *In* Proceedings of the Fourth International Autonomous Agents and Multi Agent Systems, pp. 519–526, Utrecht, the Netherlands.

MUI, L., M. MOHTASHEMI, and A. HALBERATADT. 2002. A computational model of trust and reputation. *In* Proceedings of the Thirty Fifth Hawaii International Conference on System Science, pp. 2431–2439, Manoa, HI.

PAOLUCCI, M., K. SYCARA, T. NISHIMURA, and N. SRINIVASAN. 2002. Toward a semantic web ecommerce. *In* Proceedings of the Sixth Conference on Business Information Systems, pp. 153–161, Colorado Springs, CO.

RICHARDSON, M., R. AGRAWAL, and P. DOMINGOS. 2003. Trust management for the semantic web. *In* Proceedings of the Second International Semantic Web Conference, pp. 351–368, Sanibel Island, FL.

RUSSELL, S., and P. NORVIG. 2002. Artificial Intelligence: A Modern Approach, 2nd ed. Prentice Hall, Englewood Cliffs, New Jersey.

SABATER, J., and C. SIERRA. 2002. Reputation and social network analysis in multi-agent systems. *In* Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 475–482, Bologna, Italy.

SABATER, J., and C. SIERRA. 2005. Review on computational trust and reputation models. Artificial Intelligence Review, **24**:33–60.

TEACY, W. T. L., J. PATEL, N. R. JENNINGS, and M. LUCK. 2005. Coping with inaccurate reputation sources: Experimental analysis of a probabilistic trust model. *In* Proceedings of the Fourth International Autonomous Agents and Multi Agent Systems, pp. 997–1004, Utrecht, the Netherlands.

ZACHARIA, G., A. MOUKAS, and P. MAES. 1999. Collaborative reputation mechanisms in electronic marketplaces. *In* Proceedings of the Thirty Second Hawaii International Conference on System Science, pp. 371–388, Maui, HI.