Available online at www.sciencedirect.com
ScienceDirect
International Journal of Approximate Reasoning
xxx (2007) xxx–xxx

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

www.elsevier.com/locate/ijar

On representation and aggregation of social evaluations in computational trust and reputation models

Jordi Sabater-Mir [a,*], Mario Paolucci [b]

[a] *IIIA, CSIC, Campus UAB, 08193 Bellaterra, Catalonia, Spain*
[b] *ISTC, CNR, Via San Martino della Battaglia 44, 00185 Roma, Italy*

Received 15 October 2006; received in revised form 22 November 2006; accepted 16 December 2006

Abstract

Interest for computational trust and reputation models is on the rise. One of the most important aspects of these models is how they deal with information received from other individuals. More generally, the critical choice is how to represent and how to aggregate social evaluations. In this article, we make an analysis of the current approaches of representation and aggregation of social evaluations under the guidelines of a set of basic requirements. Then we present two different proposals of dealing with uncertainty in the context of the *Repage* system [J. Sabater, M. Paolucci, R. Conte, Repage: Reputation and image among limited autonomous partners, Journal of Artificial Societies and Social Simulation 9 (2). URL http://jasss.soc.surrey.ac.uk/9/2/3.html], a computational module for management of reputational information based on a cognitive model of imAGE, REPutation and their interplay already developed by the authors. We finally discuss these two proposals in the context of several examples.

Keywords: Trust; Reputation; Aggregation mechanisms

* Corresponding author.
E-mail address: jsabater@iiia.csic.es (J. Sabater-Mir).

0888-613X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2006.12.013

1. Introduction

In the last years, the multi-agent systems paradigm has been one of the most prominent areas of artificial intelligence. In a multi-agent system there are artificial entities that have to cooperate, negotiate, coordinate, compete, etc. among them, as it is done by humans in human societies. It is clear that if we want virtual entities able to succeed in such an environment, these entities have to show some kind of *social intelligence*.

Among the different aspects associated to social intelligence, trust and reputation mechanisms [2–4] have received recently a lot of attention among AI researchers and specifically among researchers in the area of multi-agent systems. As an attempt to provide an electronic agent with the capability to evaluate the behavior of others in order to improve its social interactions with them, a significant number of computational trust and reputation models (CTR models from now on) have been developed [5–8,2,9].

Both trust and reputation can be considered as social evaluations [10] of a social target; thus, all CTR models need a representation of a social evaluation, able to express both communications and personal evaluations. In general, a social evaluation is defined as a belief of an evaluating agent about the target's usefulness with regard to a goal. In most real social situations, there is no way to make social evaluations precise; most of human social skills are based on imprecise and incomplete data, made even vaguer by the tendency to misrepresent frequencies.

Cognitive representations of social evaluations can be shown [1] to include three aspects. The first aspect is the type of evaluation -for example, trust or reputation- that could have different functional properties. The second aspect is the subject (or role) evaluation concerns: are we considering our target as a seller of goods or as an informant? The third aspect is the actual content of the evaluation: is John just good or very good? In this paper, the third aspect is the only aspect of the social evaluation that we are going to consider. In the following, when we refer to a social evaluation, we will implicitly refer to its value only.

Although each model has its own idiosyncrasies, all of them have to face a common problem: how to represent the content of the social evaluation, and how this information can be aggregated.

The information that a CTR model has to deal with in a typical multi-agent environment is of three types:

- *Direct information*, also called "direct experience". This is information that the agent obtains directly from the target, without intermediaries that can distort its content. This includes also the information obtained as a result of observing the environment.
- *Third-party information*, information that arrives through a third-party agent, the informant (or witness). The informant can be the generator of that information (that is, the information is the result of its direct experience) or only a transmitter. In both cases, the receiver cannot safely assume that the information be always true and accurate.
- *Meta information*. Knowledge that results from different analysis the agent performs on the direct and third-party information. A clear example is the information associated to the social relations among members of the society.

Third-party information is sensible to distortion and therefore it needs to be evaluated carefully before it can be incorporated as part of the agent beliefs and used to make deci-

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

sions. The mechanisms used to evaluate the reliability of this kind of information are based on previous direct information, meta-information and third-party information coming from different partners. These mechanisms, that in some models can be very complex, are beyond the scope of this article and we refer the reader to [2,3] and to the references to concrete models for further information on this topic.

We will assume (without lack of generality) that a piece of third-party information is a social evaluation in the shape of a message received from another agent with a content that provides information on a given target. At some point, once the reliability of the message has been evaluated, the CTR model aggregates the information to obtain a single measure on the target.

Our purpose in this article is specifically to study (i) how this third-party information can be represented and (ii) the mechanisms that can be used for its aggregation.

In Section 2 we discuss the requirements that we consider are important for representing and aggregating third-party information. In the following of the paper, we move on with a review of representative models already present in the literature (Section 3) and with a short overview of the Repage cognitive model (Section 4). Having thus set the stage, we present two possible interpretations of the representation of social evaluations in Repage and their corresponding aggregation mechanisms, namely the *probabilistic* interpretation (Section 5) and the interpretation that regards them as *strength of belief* (Section 6). After a comparison of the two approaches (Section 7), we finish with some conclusions (Section 8).

2. Requirements for the representation and aggregation of social evaluations

Different CTR models represent and aggregate third-party information in a different way. In this section we discuss some interesting issues about expressiveness and some relevant requirements for information manipulation, independently of the specific representation and aggregation mechanisms used.

2.1. Representation expressiveness

The choice for a representation of information has important consequences in terms of degree of expressiveness; any specific choice is bound to open some possibilities and to close others. A representation with a high expressiveness can add unnecessary complexity to the aggregation mechanism or can be cumbersome in its use for decision making. Therefore it is important to find an equilibrium between the possibility of representing as much situations as possible but at the same time maintaining a good degree of usability.

In our context (where the information represents the evaluation that a third-party agent is performing on a target) it is desirable to have a range of possibilities to express the goodness of that evaluation. This can be any numerical or non-numerical scale with an order relation on the set of all possible values [11]. Examples of numerical scales can be, for instance, an integer scale from 1 to 5 or a real scale from 0 to 1. Usually, numerical scales in the context of CTR models are of cardinal type, that is, it is possible to define operations other than comparison among them (for instance the distance between two values). The non-numerical scales are linguistic scales like those often used in fuzzy set theory, such as [bad, neutral, good] although not always these linguistic

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

terms are associated to fuzzy sets. For example, Abdul-Rahman and Hailes' model (see Section 3.3) uses a linguistic scale where the elements are just ordered categorical values. In addition to the actual value of the evaluation, some of the models introduce a degree of confidence in the evaluation, either as an independent value (see for example Section 3.4) or implicit in the actual value of the evaluation (see for example Section 3.5).

Consider the following examples of social evaluations, where an agent believes that the behavior of a target t is

- ... excellent, and I'm sure of it
- ... surely random[1]
- ... perhaps random, but I'm not really sure
- ... sometimes X, sometimes Y, with no other (middle) term (for example $X = \text{very\_bad}$ and $Y = \text{very\_good}$).
- ... I don't know anything about t

2.2. Aggregation requirements

Aggregation of evaluations is a central point in any model of reputation. As will be seen in Section 3, all the existing models have their own proposals for aggregation procedures. An aggregation procedure is an operator that, starting from a set of evaluations regarding the same target in the same role, builds a new evaluation that summarizes them. Depending on the model, this can be used to produce a single evaluation on the basis of several communications, several direct experiences, or a mix of the two. We will indicate the aggregation with the function $ag: E \times E \rightarrow E$, where E is the space of aggregation. A generic aggregation will be indicated by $ag(e_1, e_2, \ldots, e_n)$. In this paper, we will discuss several proposal for the implementation of the aggregation operator. All implementations, however, should respect the properties announced in the following paragraphs.

2.2.1. Uncertainty management

For social systems, there are several possible dimensions of uncertainty. They include simple noise, mistakes or intrinsic limitations in observation, mistakes in communications, and, even more interestingly, planned frauds and cheats, both at the pragmatic level and at the communication level [4]. While we do not plan to express that much, we think that the CTR should be able, at least, to express separately the following two situations:

- the agent thinks that the behavior of the target is unpredictable,
- the agent thinks that what it believes about the target is not certain, not dependable.

Of course, these characteristics must be conserved by the operation of aggregation so, for example, aggregating two unpredictable evaluations should result in an unpredictable evaluation again.

[1] In this context, we use the term *random behavior* to indicate a uniform probability distribution among the different possible behaviors; in other terms, the behavior is impossible to be predicted.

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

2.2.2. Consistency requirements

A plausible aggregation mechanism should show the following properties:

- *Sensibility*: the aggregation should not show large changes or jumps for small changes in the aggregating evaluations.
- *Monotonicity*: the aggregation should respect regularities in the order of the values for the contributing evaluations; for example, if the "very bad" value is lower than the "bad" value for all addenda, then this should be preserved in the result.
- *Commutativity* and *associativity*: aggregating evaluations should not depend from the order in which they are aggregated.

The last property has an obvious exception if the model takes into account time explicitly, for example discounting past experiences in favor of more recent ones. Associativity is essentially a nice property to have from a computational point of view.

In addition, to be consistent with theories of trust and reputation, the following requests should be respected:

- When adding n equal shapes, the result should be the same shape (idempotency) with a higher confidence (strength) on it.
- When adding two very different shapes, the result should be mix of them, with a lower confidence.

3. Representation and aggregation mechanisms of third-party information in CTR models, an overview

Before presenting our own proposal (see Sections 4–6), in this section we are going to summarize the main approaches on third-party message content representation and aggregation mechanisms that can be found in current CTR models. For this analysis we have selected a set of illustrative models, each one representing a different approach.

As we will see, the nature of the information that is aggregated is not always the same in the different analyzed models. In the case of the eBay and Schillo et al. models, the exchanged information are direct experiences of the informers. This means that in these cases the exchanged information (the same that later will be aggregated) is not an evaluation of the target but an evaluation of a concrete event with that target. In the rest of models the information is a reputation (that is, what the people say) or a trust degree (what the informer thinks). Either if it is a reputation or a trust degree, the information refers to an evaluation of the target and not only to concrete events.

For each model, after a short introduction we will describe:

- *Content information.* The meaning the model gives to the information received from third-party agents and the format used to represent that information.
- *Aggregation mechanism.* Algorithm used to aggregate the content information and to obtain a single measure.
- *Output format.* The format of the result of the aggregation. This format usually coincides with the format of the information that is being aggregated but not always.

Finally, in Section 3.7 we analyze the models in terms of the requirements discussed in Section 2.

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

3.1. eBay [12]

eBay is the most known marketplace in Internet nowadays. With millions of users it has become a place where it is possible to sell and buy almost everything. Among their facilities they incorporate a reputation mechanism that helps users to decide who can be a good partner. The model used by eBay is the paradigm of the models used in on-line marketplaces. Other examples are Amazon auctions [13] and BizRate [14] (the latest being a little bit more sophisticated but following the same principles).

Content information: Is the rating of an interaction a third-party (and unknown) user had with the target in the past. After an interaction, both the seller and the buyer can send a rating to the system that states their degree of satisfaction with the interaction. A rating is represented as an integer that can have three possible values: positive (+1), neutral (0) or negative (−1). For instance, a buyer that after buying a TV receives the TV at home and the TV is broken will send a negative rating (−1) for the seller. Similarly, a seller that receives late the money of a purchase can rate the buyer also with a negative rating.

Aggregation mechanism: Is a simple summation of the numerical values associated to the ratings in a fixed temporal frame.

Output format: If the summation of feedbacks is greater than 10, the target is assigned a color star that changes its color according to this number: yellow from 10 to 49, blue from 50 to 99... and similarly for the 10 possible distinctions (the maximum category achieved with a value of 100,000 or more).

3.2. Schillo et al. [6]

The model of Schillo et al. uses a probabilistic approach to calculate the final value and a boolean representation for the observed/experienced events (that is, the result of an experience can be only good or bad). The aggregation mechanism in this case is oriented to obtain a sequence of events that can be used for the probabilistic mechanism to predict what will be the next behavior of that target. This boolean perspective of the evaluations and the probabilistic approach is also shared by the model of Sen and Sajja [15] and that of Mui et al. [16].

Content information: The information exchanged by agents is an array of interaction/observation results. Each position of the array identifies a single event. If an event corresponds to an interaction between the informant and the target or has been observed by the informant, it is marked by the informant with the result of that interaction/observation that can be positive or negative (boolean value). For example, the array of interaction/observation results $[X, X, X, 1, X, 1]$ (with the array index going from 1 to 6) indicates that the informant has information about events 4 and 6 and that in both cases the information is positive.

Aggregation mechanism: "A witness (informant) will neglect only positive information but not tell something that is not the truth". This assumption is essential to be able to apply the proposed aggregation mechanism. The assumption is based in the fact that the exchanged information is a set of observed experiences (and not a summary of them). Given that, the authors assume that it is not worth it for witnesses to give *false* information. A witness will not say that a target agent has played dishonest in game x if this was not the case because the inquirer could have observed the same game and, therefore, notice that the witness is lying. Witnesses do not want to be uncovered by obviously betraying.

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

Therefore, the model assumes that witnesses never lie but that can hide (positive) information in order to make other agents appear less trustworthy. Assuming that negative information will be always reported by witnesses, the problem is reduced to know to what extend those witnesses have biased the reported data (hiding positive observations).

The first step of the aggregation process is to estimate the amount of hidden information. As we said, the witnesses never give false information but can hide positive information to bias the result, given that we know the hidden information is only positive. The calculation is based on the formula of binomial distributions and an estimation of the probability used by the witness to hide information (that can be calculated by comparing previous information of the witness with direct experiences).

Once estimated the amount of hidden information for each witness, the arrays received from the witnesses are "reconstructed" assuming that the distribution of hidden information will follow the distribution of known information.

The last step is to merge these arrays. The objective of this operation is to avoid the "correlated evidence" problem [17] (overlapped data from two or more witnesses that refers to the same event). Given the previous assumption, it is not necessary to face the problem that different witnesses give different opinions for a specific event. If they give an opinion it will be the truth and therefore must coincide with what the others (that behave the same way) say.

Once the information has been aggregated, the summary array is used to calculate the probability that the next interaction with that target be positive ($p = \#positive/\#events$).

Output format: A real value $\in [0.1]$.

3.3. Abdul-Rahman and Hailes [7]

This model uses linguistic labels to represent an evaluation. These linguistic labels, however, are not associated to fuzzy sets. The model is oriented to the recommendation of products and therefore the information exchanged is a recommendation on a specific object.

Content information: A recommended trust degree on the target. This degree is a member of the ordered set $E = \{vg, g, b, vb\}$, representing 'very good', 'good', 'bad' and 'very bad' respectively. In this case the information refers to the global opinion that the informer has on the target (and not to concrete events like in the previous two models).

Aggregation mechanism: First, the information coming from unknown agents is discarded. An agent is considered to be known when there has been a previous situation that allowed (i) to compare the provided information and the own perception of a specific outcome and from that comparison (ii) to establish a *semantic distance* to be used in future recommendations. A *semantic distance* is a measure that allows an agent to adapt the information coming from a concrete informer to its own perception. Basically it measures the deviation of the received information and the personal experience.

Second, each known informer k is given a weight according to its trustworthiness (rtd_k) and Table 1. The rtd_k value is based on the distances between the agent's recommendations and the actual experiences from relying on these recommendations.

Third, using the *semantic distance*, the information of each informer is adjusted. For instance, if the information is b and the *semantic distance* is -1 the adjusted recommendation will be vb.

Finally, for each recommended trust degree $e \in \{vg, g, b, vb\}$, sum the weights of the informers that recommended that degree. At the end we obtain a tuple of values that

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

Table 1
Informer weights in Rahman–Hailes' model

rtd_k	0	1	2	3	unknown
Weight	9	5	3	1	0

represent the 'strength' of each degree. The final combined trust degree is equal to the degree with the highest strength. If there is more than one maximum strength, the combined trust degree is assigned an uncertainty value that reflects:

- "mostly good (mg)". The maximum strengths are in the good and very good degrees with lower strengths in the bad and very bad degrees.
- "mostly bad (mb)". The maximum strengths are in the bad and very bad degrees, with lower strengths in the good and very good degrees.
- "equal amount of good and bad (gb)". All other combinations not covered till now (for example maximum strengths in the good and bad degrees).

Output format: a single value (represented by a linguistic label) $\in \{vg, g, b, vb\} \cup \{mg, mb, gb\}$.

3.4. ReGreT [18,19]

Content information: A tuple of real values $\langle Trust_{w\to t}(\varphi), TrustRL_{w\to t}(\varphi)\rangle$, where the first element $Trust_{w\to t}(\varphi) \in [-1, 1]$ is the trust value on the target (t) for a specific behavioral aspect (φ) from the point of view of the witness (w), and the second element $TrustRL_{w\to t}(\varphi) \in [0, 1]$ is a value that reflects how reliable the witness thinks the trust value is. Again in this case we are talking about a general opinion on a concrete behavioral aspect of the target.

Aggregation mechanism: The ReGreT model has a complex mechanism to evaluate the credibility of the witnesses based on social relations. Once this credibility has been established the aggregated value (a 'reputation' (R) using ReGreT's terminology) is calculated as the weighted sum of the received *Trust* values. The weight for each *Trust* value is the normalized credibility of the witness that sent it.

To calculate the reliability of a witness opinion, the agent uses the minimum between the witness credibility and the reliability value that the witness itself provides. If the witness is a trusty agent, the agent can use the reliability value the witness has proposed. If not, the agent will use the credibility of the witness (*witnessCr*) as a measure for the reliability of the information.

Given that, the formulas for the aggregation are:

$$R_{a\to t}(\varphi) = \sum_{w_i \in \mathbf{W}} \omega^{w_i t} \cdot Trust_{w_i \to t}(\varphi)$$

$$RL_{a\to t}(\varphi) = \sum_{w_i \in \mathbf{W}} \omega^{w_i t} \cdot \min(witnessCr(a, w_i, t), TrustRL_{w_i \to t}(\varphi))$$

where $\omega^{w_i t} = \frac{witnessCr(a,w_i,t)}{\sum_{w_j \in \mathbf{sW}} witnessCr(a,w_j,t)}$ and a is the evaluator that is making the aggregation.

Output format: A tuple of real values representing the reputation value and the confidence the agent has on it ($\langle R_{a\to t}(\varphi), RL_{a\to t}(\varphi)\rangle$).

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

3.5. Afras [8]

This model, like that of Ramchurn et al. (see Section 3.6) belongs to the group of models that are based on the fuzzy set theory.

Content information: A fuzzy set in a continuous space $[0, 100]$ that represent the witness reputation on a target. The reliability that the witness has on that reputation is implicit in the shape of the fuzzy set. A wider fuzzy set means less confidence on that reputation and the other way around.

Aggregation mechanism: The last information received is aggregated to the current reputation value. For this aggregation, it is used a weighted arithmetic mean. If $R_{i-1}^{EVA \to SEL}$ is the reputation of the seller (SEL) at time $i-1$ from the perspective of the evaluator (EVA) and $R_i^{WIT \to SEL}$ the communicated reputation that the witness (WIT) has on the seller, then

$$R_i^{EVA \to SEL} = w \cdot R_i^{WIT \to SEL} + (1 - w) \cdot R_{i-1}^{EVA \to SEL}$$

with

$$w = cg\left(R_i^{EVA \to WIT}\right) \in \lfloor 0, 1/2 \rfloor$$

where *cg* is the *center of gravity* of a fuzzy set.

Output format: A fuzzy set in a continuous space [0,100].

3.6. Ramchurn et al. [2]

Although the aggregation method is quite similar to that used in Afras (a weighted mean), the way fuzzy sets are used to represent the exchanged information is quite different.

Content information: The fulfillment of a contract is measured in terms of absolute variations of utility (ΔU) between the signed contract and the enacted one. Agents share a set of linguistic labels , each label $L \in \{Bad, Average, Good\}$ modeled as a fuzzy set on the domain of utility deviations $\Delta U = [-1, 1]$ specified by a membership function $\mu_L(u) : [-1, 1] \to [0, 1]$ like shown in Fig. 1.

The exchanged information between agents is a set of three *confidence levels* (one for each fuzzy set). A *confidence level* is defined as the membership level, measured over $[0, 1]$, of the behavior of a particular agent b with respect to an issue x to a linguistic term L, noted as $C(a, b, x, L)$ where a is the evaluator. The cut of the fuzzy set defined by $C(a, b, x, L)$ represents a range (on the horizontal axis) of values, that is understood as the range of expected utility deviations at execution time on issue x by agent b. Fig. 1 shows an example of communication from an informer with *confidence levels* "Good" (0.6), "Average" (0.15) and "Bad" (0). The dotted lines in the figure show where the different *confidence levels* cut the associated fuzzy set and which is the corresponding range in the ΔU axis. The shaded region indicates the range over which the sets "Good" and "Average" intersect (notice that the *confidence level* for "Bad" is 0). The base of this shaded region is the set of expected values of ΔU in a possible contract with the target that is being evaluated.

In this case, the nature of the exchanged information is the same of the previous models (Abdul-Rahman and Hailes, ReGreT and Afras), that is, the opinion that the informer has on the target but expressed in terms of utility variations.

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

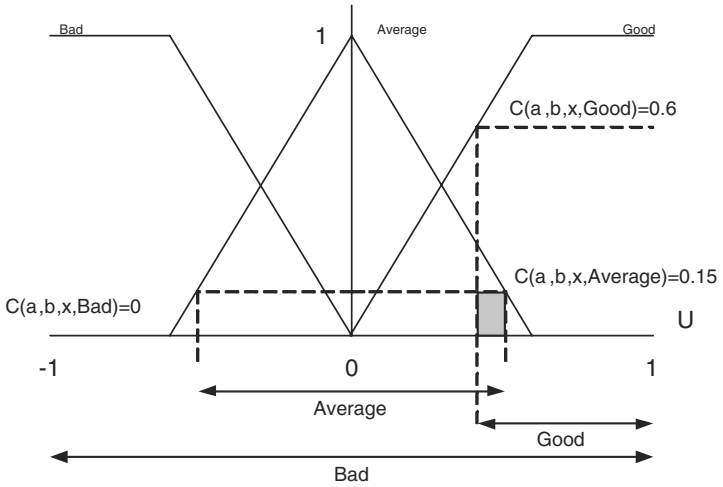

Fig. 1. Linguistic labels in Ramchurn et al.'s model.

Aggregation mechanism: This model assumes each agent belongs to a social group and that information coming from agents that belong to a relevant group (from a social perspective) has more credibility than information coming from agents in a less important group. Given that, for each linguistic label (fuzzy set) the formula to aggregate the received confidence levels is

$$Rep(a,b,x,L) = \sum_{G_i \in \mathbf{G}} w_i \times \min_{a \in G_i} C(a,b,x,L)$$

where $w_i > w_j$ if $G_i \succeq G_j$.

Output format: The result of the aggregation is a set of *confidence levels*, one for each fuzzy set.

3.7. Discussion

Expressiveness: In terms of expressiveness the representations used in Schillo et al., eBay and Abdul-Rahman and Hailes models are quite limited. They cannot express the confidence that the informant has on the transmitted value. In the case of eBay and Schillo et al. the exchanged information are evaluations of single events. In both cases the informant is sending only its direct experiences so the confidence level is not as necessary as if the information was an evaluation on the general behavior of the target like in the Abdul-Rahman and Hailes model. Of course, the fact of exchanging only single events instead of global opinions is a restriction in itself because the agent needs a lot more messages to transmit the same information. Moreover and perhaps even more relevant, giving detailed information about the direct experiences can be dangerous for the informant in environments that are not fully cooperative and friendly.

Afras has a highly expressive representation using fuzzy sets. However, by mixing the evaluation value with the uncertainty (confidence) associated to it in the same fuzzy representation, you cannot make explicit the differentiation between uncertainty and unpredictability. That is, if the agent receives a flat fuzzy set, which should be the right

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

interpretation? Does it mean the informant has no idea about the behavior of the target (uncertainty) or that the target has a chaotic (unpredictable) behavior? Ramchurn et al. model suffers from the same problem, the range in the expected utility reflects also the uncertainty so there is no way to distinguish between an evaluation that states a big uncertainty from another that is representing a great variability in the expected utility.

As we have seen, ReGreT allows the representation of the confidence on the transmitted value but the use of a simple real value to represent the evaluation limits the expressiveness of the transmitted information. For instance, there is no way to represent an unpredictable behavior.

Finally, none of the discussed representations is able to express things like "the behavior of the target is very good or very bad, no middle term".

3.7.1. Aggregation mechanisms

The aggregation mechanisms are mainly based on means and weighted means. One of the main problem of these aggregation mechanisms is how they deal with the aggregation of extreme values. A very good evaluation aggregated with a very bad evaluation results on a neutral evaluation. Those models that consider the confidence on the value can differentiate between a real neutral evaluation or an evaluation that is the result of contradictory information. This is the case for example of ReGreT where a neutral evaluation that is the result of contradictory information will have a low reliability value associated. However this solves the problem only partially because then there is no way to differentiate that evaluation from a neutral evaluation that has a low credibility because there is few evidence.

The models that aggregate direct experiences (for instance Schillo et al.) do not suffer from this problem because the exchanged information refers always to a single event but as we said, the same nature of that information restricts the use of these models to very concrete environments. Another aspect to consider is the amount of information used in the aggregation. An agent can accumulate a lot of information and the excess of it (specially if we are talking about old information) can add noise to the final result. Almost all the models consider a time frame and only the most recent pieces of information are taken into account.

Afras aggregates the most recent information with the result of previous aggregations. This has the advantage that the model do not need to store all the previous information but only the most recent aggregation and the last information received. This, however, do not allow the revision of the previous information relevance. It is normal that an informant that was considered very good in the past, now is discovered it was lying and vice versa. Given that, we consider it is worth it to store the information and be able to recalculate things considering the current knowledge on the informants.

4. The Repage model in short

In this section, we will briefly describe Repage (for a complete description please refer to [1]), a computational module for management of reputational information in an intelligent agent. Repage is based on a model of imAGE, REPutation and their interplay. Although both are social evaluations, image and reputation are distinct objects. Image is an evaluative belief [20]; it tells that the target is "good" or "bad" with respect to a norm, a standard, or a skill. Reputation is a belief about the existence of a communicated evaluation. Consequently, to assume that a target t is assigned a given reputation implies

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

only to assume that *t* is reputed to be "good" or "bad", i.e., that this evaluation circulates, but it does not imply to share the evaluation. Repage provides evaluations on potential partners and is fed with information from others and outcomes from direct experience.

To select good partners, agents need to form and update own social evaluations; hence, they must exchange evaluations with one another. If agents transmit only believed image, the circulation of social knowledge would be bound to stop soon. But in order to preserve their autonomy, agents need to *decide* whether to share or not others' evaluations of a given target. If agents transmit others' evaluations as if these evaluations were their own, without the possibility of choosing if they want to mix both types of evaluations or not, they would be no more autonomous. Hence, they must

- form both evaluations (image) and meta-evaluations (reputation), keeping distinct the representation of own and others' evaluations, before
- deciding whether or not to integrate reputation with their own image of a target.

Unlike current systems, given what we have said above, in Repage reputation does not coincide with image. Indeed, others can either transmit their own image of a given target, which they hold to be true, or report on what they have "heard" about the target, i.e., its reputation, whether they believe this to be true or not. Of course, in the latter case, they will neither commit to the information truth value nor feel responsible for its consequences. Consequently, agents are expected to transmit uncertain information, and a given positive or negative reputation may circulate over a population of agents even if its content is not actually shared by the majority.

The main element of the Repage architecture is the memory that is composed by a set of *predicates*. Predicates are objects containing a social evaluation, belonging to one of the main types accepted by Repage (image, reputation, shared voice, shared evaluation), or to one of the types used for their calculation (valued information, evaluation related from informers, and outcomes). As we will see later in more detail, these predicates have a tuple of five numbers to represent the evaluation plus a strength value that indicates the confidence the agent has on this evaluation. Predicates are conceptually organized in different levels and interconnected to reflect their dependencies. To have a pictorial summary of the kinds of predicates in the Repage's memory, and of their relative position, refer to Fig. 2; for a detailed description, confront [1].

Predicates are connected by a network of dependencies, that specifies which predicates contribute to the values of other ones. Each predicate in the *Repage* memory has a set of antecedents and a set of consequents. If an antecedent is created, removed, or changes its value, the predicate is notified, recalculates its value and notifies the change to its consequents.

This is not the place to proceed with a detailed explanation of all the mechanisms in Repage, for which we refer again to [1]. The point here is that in order to make use of Repage's complex network of dependencies we will need dependable algorithms for the aggregation of social evaluations. These algorithms must be designed by taking into account what exactly is expressed by the chosen representation for social evaluations. As will be shown in the following, even with the same mathematical structure, differences in interpretation can bring about very different choices for the manipulation of social evaluations. But before entering in the details of the aggregation choices, we will explain what is the mathematical structure inside Repage's predicates and what is its design rationale.

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

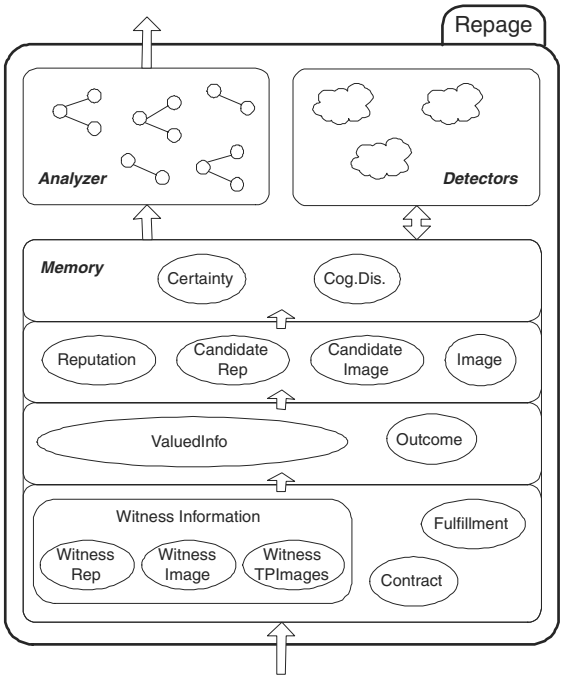

Fig. 2. *Repage* architecture as a module for an intelligent agent. In the Memory, several kinds of predicates are organized in different levels. The detectors build new predicates on the basis of the existing ones. The analyzer give indications to the rest of the system.

4.1. On representations and uncertainty in Repage

Agents must keep track of social evaluations. This means that they need the content of the evaluations – in simple systems, a position on a scale going from good to bad. Thus, we could simply associate to a predicate a scalar number.

An interesting question arise when we wonder what is the correspondent of the intuitive concept of an uncertain evaluation. For a number on a scale, the only available representation is the middle of the scale. But this representation will have multiple semantics, and the uncertainity will get mixed with actual regular middle outcomes.

An alternative can be provided by the usage of a *labeled tuple*, that is, a tuple of five numbers each of which has an associated label in a rating scale, for example from ''very bad'' to ''very good''. A labeled tuple is more expressive and allow us two dimensions of uncertainity – that is, a ''spike'' in the middle of the scale or a ''flat'' evaluation. This expressiveness allows us to distinguish a target that is constantly mediocre (the spike) from one that is actually unpredictable or uncertain (the neutral or flat evaluation).

There are several contexts where this is important. Consider for example one in which some of the classes are extremely undesirable; in the case of evaluations about the respect of a norm, an agent could have a goal to avoid altogether the possibility of meeting with ''very bad'' agents – in this case, it would be important to distinguish the flat evaluation, that does not deny the possibility for a''very bad'' encounter, from the spike in the middle, that excludes it. The same kind of considerations could be done for the case in which the middle value is actually important.

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

Thus, we argue that oversimplistic representations of social evaluations – of the kind a number on a scale – are insufficient to represent situations that are cognitively plausible. However, not even the labeled representation proposed until now is suitable as the representation of a social evaluation – the predicate content – in a model of a cognitive and social mind as Repage is.

There is another ingredient that must be added, that is, the strength of belief of the agent holding the representation in the representation itself. This last ingredient is indispensable to cover for yet another level of uncertainty. In fact, no model of reputation is worthy of its name if it does not accounts for the social transmission of evaluations.

But to consider reported information requires the ability to accept information "with reserve". To this end, we added a scalar value as the strength of belief that the holder has in the evaluation, similar to the one in the ReGreT [19] model. This strength value will drive the aggregation process; an evaluation with a large strength represent a certain belief, that will preserve its shape (intended as the relation between weights) if combined with another, less strong evaluation. As an example, consider an aggregation of many evaluations, composed by a large majority that shows a consensus (very similar or identical shape) plus some outliers; all of them with the same, not very large, strength. The aggregation of the similar group will produce a very strong evaluation. When adding the outliers, the strong shape will remain more or less unchanged.

In Repage, this strength value is a function of (i) the strength of its antecedents and of (ii) some special characteristics intrinsic to that type of predicate. For instance, the strength of an *Image* is a function of the strengths of the antecedents (outcomes, information from third-party agents and their image or reputation as witnesses, ...) but also of the number of these antecedents.

We maintain the content of a predicate as a tuple of five numbers (summing to one) plus a strength value. Each number has an associated label in the rating scale: very bad (vb), bad (b), neutral (n), good (g) and very good (vg). We call this representation a *weighted labeled tuple*.

In mathematical terms, we represent this tuple as $[w_{vb}, w_b, w_n, w_g, w_{vg}]$, or , for short, $[w_1, w_2, \ldots, w_5]$, where w_1 corresponds to very bad and w_5 to very good. The sum of the five components is fixed to 1. In addition, we have a single value indicating the strength of belief in the evaluation, a number $s \in [0, 1]$. In the following, we will express this evaluation as the tuple $\{[w_1, w_2, \ldots, w_5], s\}$. We will sometime refer to the set of the weights as the *shape* of the weighted labeled tuple.

In Fig. 3, we show some examples. Example (a) shows an evaluation that says the agent is 0.8 sure (almost completely certain) that the behavior of the agent evaluated is usually very bad or sometimes bad. In (b) the evaluation describes a behavior that is always very good or very bad (a black or white behavior with no grey) although the evaluator is quite unsure about it ($s = 0.2$). In (c) the evaluator is saying it is completely certain ($s = 1.0$) that the behavior of the agent is unpredictable. Notice the difference between (c) and a situation where the strength is 0. When the evaluation has a strength of 0, the evaluator is saying it does not know anything about the agent that is being evaluated and that the shape of the membership is meaningless.

The expressiveness of this representation is clear in examples like (b) or (c) that would be impossible to reproduce using two single real values (like, for instance, in the ReGreT model; see Section 3). Intuitively, this choice for the predicate adds more levels of freedoms and allows for subtler representations. But although the examples presented seem to make perfect sense by themselves, they actually leave space for different interpretations.

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

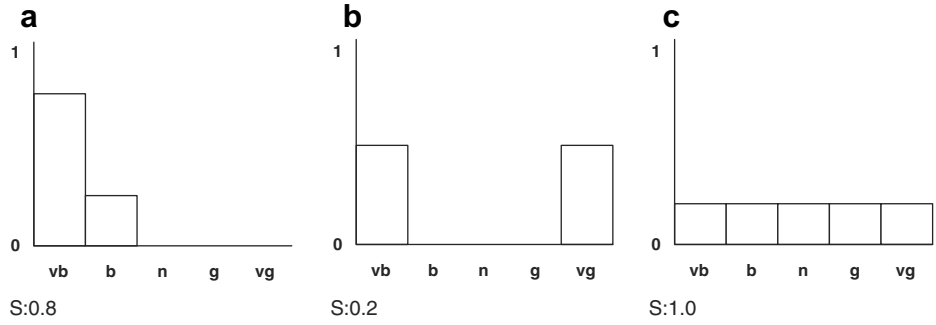

Fig. 3. Examples of evaluations in Repage.

Are the weights probability to obtain the result corresponding to their label? Or do they express simply a possibility? There is a large body of literature about the subtelties in the representations of uncertainty; see for example [21], where the author distinguishes between probability measure, possibility measure, and belief measures; our approach falls in this last area, even if it is much more practical and not based on a possible worlds logic. Indeed, we are particularly interested in the case in which the weights express the strength of belief of a decision maker in the corresponding outcome, as proposed and discussed in [22], that refers to the general area of decision making under uncertainty (see also [23]). Even if these points of view seem very similar, when used as a base for designing the aggregation algorithm they appear to be very different.

In the first case, the *probabilistic* approach allows the aggregation method to be simple and not demanding from a computational perspective. To the contrary, the second approach, that we will call the *strength of belief* one, will be the base for more complex considerations and a large amount of subtleties, resulting in an improved expressive power at the expenses of computational simplicity.

5. Probabilistic approach

In this section we present what we call the probabilistic approach. First we will discuss the interpretation we give to the representation and its properties and then we will detail the aggregation algorithm proposed.

5.1. Representation

In the probabilistic approach, the weights $w_i \in [0,1]$ in the labeled tuple represent the probability that the evaluation of the future behavior of the target agent will be classifiable under the labels $i \in \{vb, b, n, g, vg\}$. Because w_i are probabilities, we have that $\sum_{\forall i} w_i = 1$. The strength value $s \in [0,1]$ associated to this tuple represents the degree of certainty of the probability distribution represented in the tuple. We remember the representation for the weighted labeled tuple as $\{[w_{vb}, w_b, w_n, w_g, w_{vg}], s\}$.

Notice that the "shape" of an evaluation do not represent uncertainty. In this approach, the uncertainty is defined as the lack of knowledge about the behavior of an individual. Therefore if you have a "flat" shape (same probability for all the possible behaviors) you actually know which is the expected behavior (random in this case) and

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

therefore there is no uncertainty. The uncertainty is reflected only by the strength value. Given that, it is wrong to assume that the uncertainty reflected by the strength can be represented as a probability uniformly distributed among all possible behaviors. If that was the case, the strength value would not be necessary anymore. However, talking about behaviors you cannot assume that the uncertainty has a uniform distribution over the possible values (like when you throw a dice, for example).

5.2. Aggregation algorithm

In this approach, there are two things that have to be taken into account: how to aggregate the probability distributions and how to aggregate the strengths. We will indicate the aggregation function as ag_p.

For the aggregation of the probability values we opt for a weighted mean of the probabilities associated to the same behavior using the strength of the corresponding evaluation as the weight. That is, if we are aggregating m evaluations $E = \{e_1, \ldots, e_m\}$ where $e_j = \left\{[w_{vb}^{e_j}, w_b^{e_j}, w_n^{e_j}, w_g^{e_j}, w_{vg}^{e_j}], s_j\right\}$, we obtain an evaluation $e = ag_p(e_1, e_2, \ldots, e_m)$. The weights w_i of e are calculated as

$$w_i = \sum_{j=1}^{m} \frac{s_j \cdot w_i^{e_j}}{\sum_{k=1}^{m} s_k} \quad \forall i \in \{vb, b, n, g, vg\}$$

For the aggregation of the strengths, it is important to notice that it is not possible to aggregate them without taking into account the probability distributions they are associated with. Not only that, the aggregation mechanism has to be sensible to the degree of similarity of the evaluations that are being aggregated. For instance, assuming the same strength, if the evaluations go in the same direction the resulting strength should be higher than if they are conflicting.

We propose the following mechanism for the aggregation of the strengths. For each evaluation to be aggregated we calculate a distance between the probability distribution associated to that evaluation and the resulting distribution after the aggregation of the probability values. This distance is a measure of the degree of conflict of this evaluation with the rest. A low distance value means there is a high coincidence with the rest of evaluations and the other way around.

The distance between two labeled tuples considered as probability distributions is calculated as the angular momentum of inertia of the shape resulting from the difference between the two tuples, that is, the tuple $d = \left[\left|w_1^1 - w_1^2\right|, \ldots, \left|w_n^1 - w_n^2\right|\right]$, $d_i = \left|w_i^1 - w_i^2\right|$ (note that d does not sum to one). The reason is that we want to give more relevance to differences situated in the extremes (*vg*, *vb* labels) with respect to ones in the middle. In other words, a shape that is more difficult to rotate (e.g., $[1, 0, 0, 0, 1]$) will express a larger difference than one with the same area but situated in the center (e.g., $[0, 0.2, 0.6, 0.2, 0]$). For the momentum we first calculate the center of mass

$$c = \sum_{k=1}^{n} (k-1) d_k \Big/ \sum_{k=1}^{n} d_k \tag{1}$$

If the denominator of the fraction is zero, it means that all the differences between weights are zero too; in this case the distance is zero. The angular momentum that we will use as distance is then defined by

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

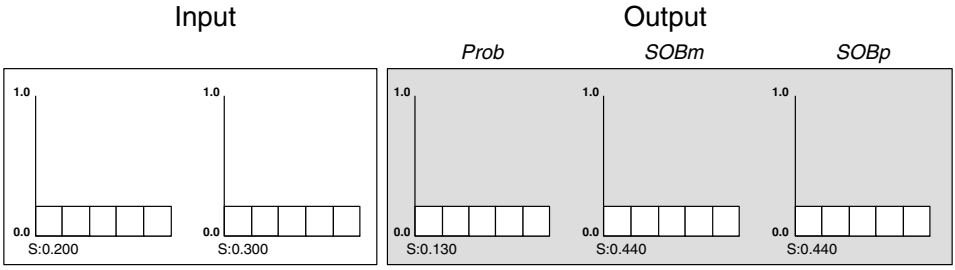

Fig. 4. Aggregation of two low-strength neutrals. Input has white background, output has gray background; in sequence, output from algorithms *Prob*, *SOBm*, *SOBp*. The shapes are maintained.

$$\mathrm{dist}(w^1, w^2) = \sum_{k=1}^{n} d_k(k - 1 - c) \tag{2}$$

The final strength is calculated using the formula

$$s = \min\left(1, \sum_{j=1}^{m}(s_j \cdot (1 - \mathrm{dist}_j))^{\rho}\right) \tag{3}$$

where s_j is the strength of evaluation e_j and $dist_j$ the distance between the probability distribution of evaluation e_j and the probability distribution after the aggregation. With the ρ parameter it is possible to diminish the relevance of low certainty values.

Notice that with $\rho > 1$ the expected behaviors discussed in Section 2.2.2 are not observed when we are aggregating a few number of low-strength evaluations (see Fig. 4 in Section 7). In a general context this is not desirable and ρ should be maintained equal to 1. However in the context of Repage there is a second possibility. As we said, Repage organize predicates in different levels. Predicates in lower levels are aggregated to get predicates in upper levels. Given that, you can consider that to really contribute to a higher level predicate it is necessary a minimum strength in the lower level predicates. Lower level predicates with a low strength are not "good enough" to contribute significantly to higher level predicates. This is what parameter ρ is intended for. With $\rho = 2$ for instance, the agent needs a lot of low-strength coincident informations to increase the certainty and only at a strength level of 0.6 the pieces of information start to contribute significantly. In the examples presented in Section 7 we will use a $\rho = 2$ to compare with the aggregation of strengths used in the *Strength of Belief* approach presented in the next section that is similar to the situation with $\rho = 1$.

Finally, remark that because we are not normalizing the aggregation of strengths, the summation of strengths after being adjusted according to the distance could be bigger than 1. That is not a problem in our context and it has complete sense. If you are *completely certain* (strength = 1) you cannot be more certain than that, so the evaluations that are rising the strength even more, only are reinforcing the completely certain feeling. This idea is represented by the minimum between 1 and the aggregation of the strengths.

6. Evaluations as strength of belief

In this section, we will propose a method to aggregate evaluations considered as Strength of Belief (SOB). Interestingly, this interpretation adds more stringent requirements to the ones already examined for the probabilistic interpretation.

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

In the following, we discuss these requirements and illustrate how our proposed algorithm is able to meet them. We will start with a section on the aggregation of the values, where we propose essentially a solution taken from the literature, and we move then to our original proposal for calculation of strengths.

There is not a single technique to aggregate tuples sets; in fact, aggregation techniques are currently debated in the fuzzy set community. We refer to the work of Yager [22,24] that present an extended discussion of the possible aggregation modalities.

That approach is connected not to fuzzy representation of *evaluations*, but to representation of alternatives in decision making under uncertainty (DMUU). The values in the labeled tuple correspond to the strength of belief that a given outcome will occur in other words, the plausibility of some alternative events. As an example, we could think of an agricultural agent examining the consequences of weather conditions; the object of interest is then composed from the set of possible weather conditions, each one connected to its utility value (considered as objective) and to the strength of belief that condition is going to occur.

However, a simple parallel with the social evaluations can be drawn. We could consider the values used in Repage as representing the strength of belief in the corresponding characterization of the target. This amounts to interpret the weights composing the labeled tuple as a measure of uncertainty about a defined (but hidden) reality – the target is actually either very bad, bad, neutral, good, or very good, but we do not know that.

6.1. The neutral value

We define as the identity value for aggregations a flat evaluation, where all values are equal (we call it "the neutral" from now on). If we are working with tuples having n components, the weights of the neutral $e_{1/n}$ will obviously have value $1/n$. Aggregations with the identity leaves evaluations unchanged: $ag(e, e_{1/n}) = ag(e_{1/n}, e) = e$.

Following [24], we defend that the use of the neutral has interesting semantic properties. First of all, it is the natural choice for SOB when we know nothing. In fact, in the interpretation of SOB as guidelines for choice, the neutral is not recommending any alternative over any other. Moreover, scores greater than $1/n$ can be considered as affirmative or expressing positive support for the related outcome. The opposite is also true; when a score is lesser than $1/n$ it represents a rejection of that particular outcome. Aggregation operations should preserve the validity of these semantic properties. Thus, when aggregating two values

- if both values are under/over $1/n$, the aggregation should result respectively in a lower/higher result value, reinforcing the tendency signaled by the individual operands.
- when adding two evaluations with the same relative ordering between outcomes, the result must show the same ordering as the addenda.

6.2. Aggregation of evaluations as SOB

For the aggregation of values, we refer to [24]. In particular, we will choose one strategy proposed for the generalized combination of sources with different credibility. From the large set of aggregation strategies proposed by Yager we adopt the simplest in our view, by defining an aggregation strategy that respects the neutral as unity.

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

We are interested in the aggregation operation for m evaluations, $e = ag(e_1, e_2, \ldots, e_m)$. The weights of the addenda are represented by $w_i^{e_j}$, where the lower index i refers to the different weights of the same evaluation, and the higher one j is used to distinguish the evaluations to aggregate. The strength of an evaluation j is represented by s_j. We will calculate the weights w_i and the strength s of the aggregation e.

Just before performing the aggregation we mix the two kinds of uncertainty – the scalar strength s and the distance from the neutral – by rescaling each addendum toward the neutral. In formulas, we rescale each weight w_i to $w_i^* = w_i s + (1 - s)\frac{1}{n}$.

To express the calculation strategy for the aggregation of m weights $Agg(w_1^*, \ldots, w_m^*)$, we drop the weights equal to $\frac{1}{n}$ and divide the set of remaining weights in two subsets L and H, respectively with value lower/higher than $\frac{1}{n}$; the cardinality of the sets is respectively n_L, n_H; $n_T = n_L + n_H$. We define

$$Agg(w_1^*, \ldots, w_n^*) = \frac{n_L}{n_T} T^*(L) + \left(1 - \frac{n_L}{n_T}\right) S^*(H)$$

Note that if $n_T = 0, n_L$ is zero too, and this means that all aggregating weights are equal to $\frac{1}{n}$. In this case, the result is again $\frac{1}{n}$.

We still have to choose the functions T^* and S^*; a discussion of their properties is beyond the scope of this paper and we refer the reader to the many possible choices in [24]. In this paper, we will examine two possible solutions. The first is the *max–min* approach, with $T^*(l_1, \ldots, l_k) = \min_{j:1\ldots k}[l_j]$ and $S^*(h_1, \ldots, h_m) = \max_{j:1\ldots m}[h_j]$.

The second is the *probabilistic sum*, where we define

$$T^*(l_1, l_2) = n l_1 l_2, \quad T^*(l_1, l_2, l_3) = T^*(l_1, T^*(l_2, l_3))$$

and so on; For S^* we give a recursive definition, starting from

$$S^*(h_1, h_2) = \frac{h_1 + h_2 - h_1 h_2 - \frac{1}{n}}{1 - \frac{1}{n}}$$

Terms with more variables are defined by $S^*(h_1, h_2, h_3) = S^*(h_1, S^*(h_2, h_3))$, and so on. Once defined the aggregation function, a complete aggregation strategy will follow the routine:

(1) rescale each tuple toward the neutral as a function of strength;
(2) calculate the tuple resulting from aggregation. The result depends from the choice for S^* and T^* – in this paper, we propose to choose between the max–min and the *probabilistic sum* formulas.
(3) calculate the strength of the result, based on a measure of difference between the result itself and the contributions. This calculation will be examined in the following section.

6.3. Calculating strength

The requirements for the calculation of strength are that the value should be between 0 and 1; when adding very similar evaluations the result should be of the same shape with higher strength, and that when adding two very different things the result exhibits lower strength.

A possible set of functions whose characteristics can be useful for positive strength calculations is

$$r(x, y) = x^p + y^p - (xy)^p \tag{4}$$

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

varying the p parameter between zero and infinity. These functions, for $x, y \in [0,1]$, are limited between zero and one and growing in both x and y. Moreover, $r(x,y) \geqslant x, y$; $r = 1$ if x or y is 1, and if one of the values is zero r is equal to the other one. The exponent p controls the rate of growth between zero and one; in the present work, we will limit ourselves to the exponent $p = 1$. In that case, once fixed one of the arguments, the function is linear in the other one. We note also that the function can be recursively generalized to $r(x,y,z) = r(x, r(y,z))$ and so on. To use this function in the calculation of strength, we need now a way to calculate the difference between the shapes of the evaluations.

6.3.1. Factors in the difference between two shapes

We will here propose a method to calculate the difference in shape between two evaluations. Since we are only interested in shape, we will operate on the weights only; we will denote them as η and θ, whose weights are respectively w_i^η and w_i^θ. In terms of evaluations, of course $e = \{\eta, s\}$. Unlike real numbers, there is a large arbitrariness in the metrics that can be used to calculate the difference between two shapes. A reasonable procedure must produce a value that is very low when comparing similar evaluations, and high when the evaluations are incompatible or very different.

A starting point is the calculation of the total absolute difference in area from the two evaluations, that is, $tab(\eta,\theta) = \sum_{i=1}^{n} \left| w_i^\eta - w_i^\theta \right|$. This value is bounded between zero (for identical evaluations) and two (for incompatible evaluations).

However, this brings the same result for the following two couples: $[1,0,0,0,0]$, $[0,0,0,0,1]$ and $[0,0,1,0,0]$, $[0,0,0,1,0]$. Indeed, we have ignored the ordering existing between the different outcomes, from very bad to bad.

To take into account this final ingredient, we resort to a variation on the technique proposed for [1], calculating a kind of difference between the centers of mass of the evaluations, that is, $cm(\eta) = \frac{\sum_{i=0}^{n-1} i \cdot w_i^\eta}{n-1}$. This amounts to $dcm(\eta,\theta) = |cm(\eta) - cm(\theta)|$ that is bounded between zero and one.

6.3.2. Calculating the difference between two evaluations

The elements shown above (*tab* and *dcm*) are all relevant contributions for the calculation of the aggregated strength. We must now decide how to combine them. There are several workable solutions to this effect; any function $f : [0,1] \times [0,1] \to [0,1]$, not decreasing in all arguments, will do. The first obvious candidate – multiplication – will have the undesirable effect to bring zero whenever one of them is zero, even if the others are large. This makes sense for *tab*, that is zero only when two evaluations are identical. Much less so for *dcm* – for example, $[0,0,1,0,0]$ and $[0.5,0,0,0,0.5]$ have the same center of mass. To avoid this we move the contribution of this term between one and two, as in the following formula:

$$df(\eta,\theta) = \frac{1}{2} tab(\eta,\theta)(dcm(\eta,\theta) + 1) \tag{5}$$

This value should be used in the calculation of strength. When $df = 1$ the resulting strength should be much lesser than the contributing ones, and when $df = 0$ we should have the maximum of reinforcement, as we proposed, the r function. In other words, for the calculation of the strength s of the evaluation $ag(e,f)$, if $e = \{\eta, s^e\}$ and $f = \{\theta, s^f\}$,

$$s^{ag(e,f)} = (1 - df(\eta,\theta)) r(s^e, s^f)$$

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

6.3.3. Calculation of resulting strength

We have now all the ingredients needed to compute the strength relative to an aggregation of m weighted labeled tuples $e_1, \ldots, e_m$; e_i can also be written as $\{\eta_i, s_i\}$ to separate shape from strength. We will calculate the resultant from the contributing strengths, divided between supporting evidence and contradicting evidence. We mark this distinction, somehow arbitrarily, to a value of difference of one half.

After calculating the rescaled evaluations $e_1^*, \ldots, e_m^*$ and the resultant shape as the labeled tuple $v = [Agg(w_1^{e_1^*}, \ldots, w_1^{e_m^*}), \ldots, Agg(w_n^{e_1^*}, \ldots, w_n^{e_m^*})]$, we calculate the set of differences $\{df(v, \eta_1^*), \ldots, df(v, \eta_n^*)\}$. This set is divided in two parts, the first set $S = \{k : df(v, \eta_k^*) < \frac{1}{2}\}$ including all supporting evidences and the second set $C = \{k : df(v, \eta_k^*) > \frac{1}{2}\}$ containing the contradicting evidences.

The strengths relative to both sets are then rescaled in function of the distance from the result, obtaining the two sets $Ss = \{s_k \cdot (1 - 2df(v, \eta_k^*))\}_{k \in S}$ and $Cs = \{s_k \cdot (2df(v, \eta_k^*) - 1)\}_{k \in C}$. The first set will build the positive part of the strength, $s^+ = r(Ss)$. The second set will build the negative part of the strength, using the same formula: $s^- = r(Cs)$.

To obtain the value for the aggregated strength we simply take the difference between s^+ and s^-, when positive, or zero if negative: $s = \max(s^+ - s^-, 0)$.

7. Comparison between the proposed approaches

To compare the different approaches, we will check them against the requests made in Section 2. In the following, we will refer to the probabilistic approach (Section 5) as *Prob*, and to the approach based on strengths of belief (Section 6) as *SOB*. When the distinction between the min–max and the probabilistic sum choices is important, we will refer to them as *SOBm* and *SOBp*. For the aggregation of strengths in the *Prob* approach we will use $\rho = 2$ as explained in Section 5.2.

To begin, let us first check how we would express the examples presented in 2 with the proposed formalism. In the examples, an agent believes that the behavior of a target t is

- ... excellent, and I'm sure of it: {[0, 0, 0, 0, 1], 1}
- ... surely random: {[0.2, 0.2, 0.2, 0.2, 0.2], 1}
- ... perhaps random, but I'm not really sure: {[0.2, 0.2, 0.2, 0.2, 0.2], 0.4}
- ... sometimes very bad, sometimes very good, with no other (middle) term: {[0.5, 0, 0, 0, 0.5], 1}
- ...I don't know anything about t: any tuple with strength zero.

About uncertainty, in both cases we can express unpredictability (a flat shape) and uncertainty (low strength). However, the interpretations are slightly different: while for *Prob* we interpret the neutral as all results being equiprobable, the same object in the *SOB* interpretation means that there are no facts to prefer any of the results to the other.

We can also check what happens when we try to aggregate two neutrals in the same strategies. The results can be seen in Fig. 4 for low strengths and Fig. 5 for high strengths. Notice that in all cases, as expected, the shapes are maintained. The difference is the way the strength is aggregated. The *Prob* approach minimizes the relevance of low-strength inputs and maximizes those inputs with high strengths. This is because we are using a $\rho = 2$ in the formula for the aggregation of strengths as explained in Section 5.2.

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

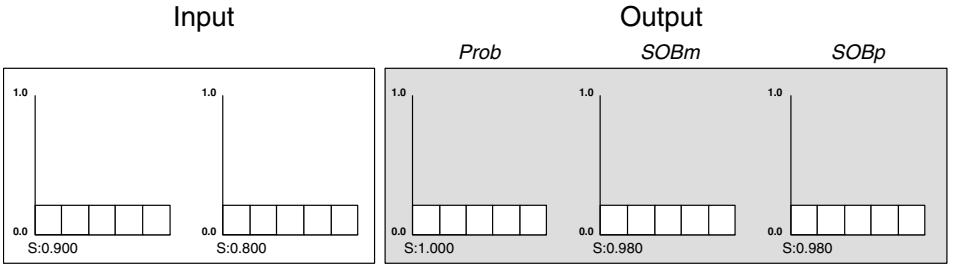

Fig. 5. Aggregation of two high-strength neutrals. Input has white background, output has gray background; in sequence, output from algorithms *Prob*, *SOBm*, *SOBp*. The shapes are maintained and strength is reinforced; note that the *Prob* approach converges to certainty of randomness, while the *SOB* approaches have a slower, asymptotic convergence.

7.1. Consistency requirements

About the consistency requirements, our proposals satisfy most of them, with some exceptions:

7.1.1. Sensibility

The algorithms are designed over continuous functions; this simple observation should be enough to hint toward their stability with respect to small changes.

To calculate the sensibility to change, the generic change of an aggregation $z = ag(e_1, e_2)$ (the generalization to more than two addenda is straightforward) can be divided in a change in strength and one in the tuples' components – the latter, of course, should be compensated to keep the sum constant. Actually, every change in component's value can be decomposed in simple changes of two components $w_i \to w_i + \epsilon, w_j \to w_j - \epsilon$, with of course $\epsilon > 0, i \neq j$; we search for the difference between z and $z' = ag(e_{1,\epsilon}, e_2)$. In the following, we will argue that this difference is limited for all proposed algorithms.

For the *Prob* approach the change in all shapes is limited by ϵ; in fact, for the changed components, if we indicate the strength of e_1, e_2 with s_1, s_2 respectively, we have that $|w_i^{z'} - w_i^z| = \frac{\epsilon s_1}{s_1+s_2} < \epsilon$. For the unchanged components, the difference is zero.

What about the change in strength of the result? This is slightly more complicated, since it takes into account the changed distance between the two evaluations. To calculate the resulting strength, as defined in Section 5.2, we will need the center of mass of the difference between components (1), the momentum of the same (2), and then their sum as for (3). Showing a detailed calculation is beyond the scope of this work; to defend the lack of jumps in the strength, we only need to observe that $|s^{z'} - s^z|$ a function of ϵ is a composed function of continuous functions,[2] and as such continuous in ϵ.

The situation is even simpler for what regards a small change in strength in the first component e_1; this will produce a change in shape, this time for all components, that at the first order in ϵ will amount to $w_i^{z'} = w_i^z + (w_i^{e_1} - w_i^z/(s^1 + s^2))\epsilon$.

[2] The only possible problem with the calculation of strength lies in the denominator of (1). But it is easily seen that if this is zero, then the two evaluations have the same shape initially, and after the perturbation they will differ by an amount proportional to ϵ. The same argument is also valid in reverse, i.e., if the perturbed components are identical, then the unperturbed ones must differ by an amount proportional to ϵ.

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

As for the SOB approach, the aggregation is designed to be continuous, as is the algorithm for strength aggregation. The only possible point of discontinuity is given by some addenda moving between the set of supporting evidence and the set of contradicting evidence. When a distance from the result passes through $\frac{1}{2}$, even if the change in shape is small, the corresponding evaluation will move from one set to the other – let us say from supporting to contradicting. If this evaluation is strong, this could cause a jump in strength. But this effect is limited by the factor $1 - 2df(v, e^*)$ that will be very small at that point.

An example for this case is presented in Figs. 6 and 7. In these figures, for the case of *SOBp*, we are passing from a distance of about 0.505 (contradicting, Fig. 6) to a distance of about 0.497 (supporting, Fig. 7). A change in strength of 0.01 is reflected by an even lower change in the result. Note also that while we are at the turning point for *SOBp*, this is not true for *SOBm*, since the results of aggregation are different in shape. The difference in strength between the two *SOB* approaches is solely due to the fact that the distances are calculated against the results, which are slightly different in shape.

7.1.2. Monotonicity

The aggregation should respect regularities in the order of the values for the contributing evaluations. Both *Prob* and *SOB* algorithms are designed exactly for this purpose.

In the case of the probabilistic approach we are using a weighted mean where the aggregated values are always positive so the monotonicity is guaranteed. For the demonstration of the monotonicity property in the *SOB* approach; we refer to the one in [24, p. 138], that applies to the more general case in which S^* is derived by any t-conorm and T^* is derived by any t-norm. An example of this property can be seen in Fig. 8.

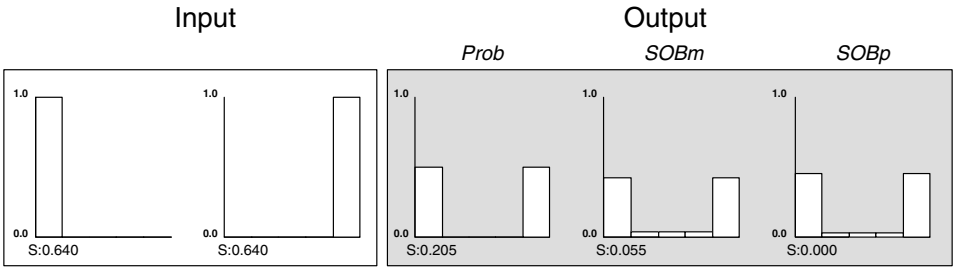

Fig. 6. Aggregation at the turning point for *SOBp*, contradicting case. Note the third result with zero strength.

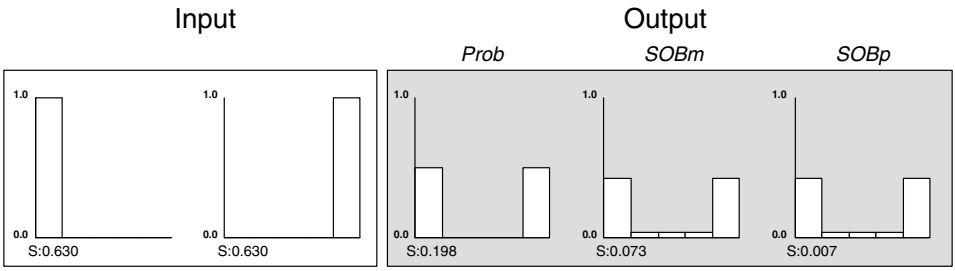

Fig. 7. Aggregation at the turning point for *SOBp*, supporting case. Compare with previous figure, third result. A small change in input strength results in a small change in result strength.

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

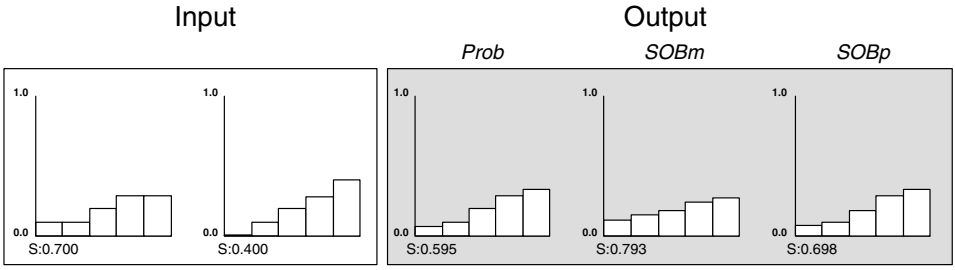

Fig. 8. Aggregation of ordered evaluations. Monotonicity is respected for all algorithms (from left to right, *Prob*, *SOBm*, *SOBp*).

7.1.3. Symmetry and associativity

Aggregating evaluations should not depend from the order in which they are aggregated. While the algorithms that we propose are not really associative, they provide an explicit recipe for calculating aggregation of any number of terms.

The aggregation algorithm in the probabilistic approach is using a weighted mean so, in a strict sense, it does not have the property of symmetry [25]. However, in our case it has no sense to dissociate the weight (the strength of the evaluation in our case) and the value (the components w_i of the evaluation). The weight is not only associated to the source, depending also on the value that is transmitted. Therefore we have to consider the tuple (weight,value) as a whole. Given that, it is clear that the operator is symmetric with respect to the tuple and the order in which the data is presented makes no difference; as in the example provided by Figs. 9 and 10. The same can be said with respect to the SOB operators, that are defined as composition of symmetric functions; the parts composing the *Agg* function, S^* and T^*, are symmetric in their arguments for both the max–min and

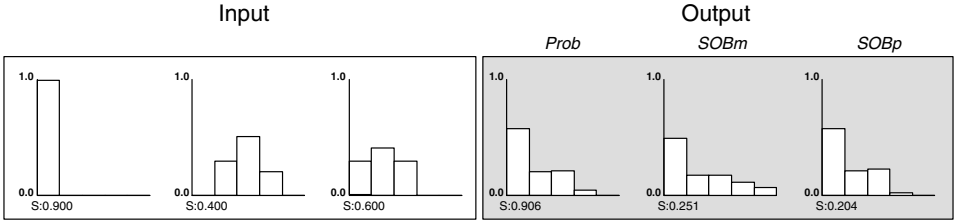

Fig. 9. Aggregation of three evaluations. Compare with next figure.

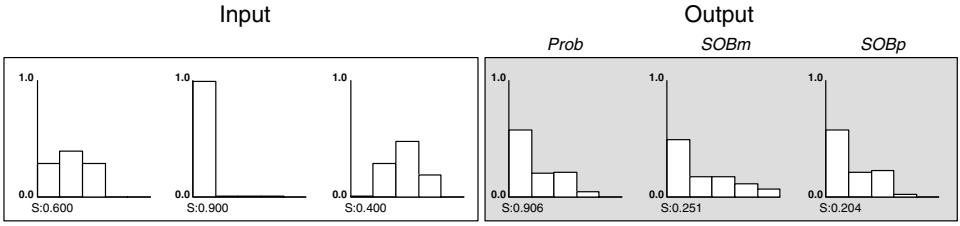

Fig. 10. Aggregation of three evaluations.

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

the probabilistic sum approaches, and the subdivision of the weights in the L and H sets is independent of the order in which the weights are presented. The aggregation is therefore symmetric for the *SOB* approach as well.

8. Conclusions

In this paper, we have reviewed current approaches to the aggregation of social evaluations, with a focus on third-party information. These systems vary from the aggregation of simple scalars to the manipulation of complex fuzzy evaluations. The decision to use the one or the other of these approaches cannot be made without reference to the specific context; what can be provided is instead a list of plausible requirements for an aggregation function, requirements that we have presented in Section 2, ranging from expressiveness (it is important to be able to express and correctly characterize uncertainty) and consistency of the aggregation operations. After examining the models present in the literature, we shortly present Repage, a module for management of reputational information, where we try to represent with some accuracy the cognitive structure of reputation. To this purpose, we need to be able to manipulate uncertain information and to accept it with reserve; the weighted labeled tuple, introduced in Section 4.1, appears to be expressive enough for this representation. However, the addition of degrees of freedoms must be done while maintaining a clear understanding of the interpretation desired for the representation of these social evaluations. Indeed, we show that different interpretations – the probabilistic interpretation and the interpretation as strength of belief – can bring about very different operators for aggregation. After presenting these operators, we discuss several examples that contribute to show similarities and differences.

While this work is inspired in our analysis for the management of third-party information in trust and reputation systems, our considerations can be extended to the more general area of social evaluations. Indeed, the algorithms presented have an use in any field related with decision making under uncertainty.

Future extensions of this work will include simulative testing of the algorithms in various contexts, from simple markets to simulated electronic auctions to social network analysis. The authors are currently deploying such simulations, integrating in the Repage system the techniques presented.

Acknowledgements

This work was partially supported by the European Community under the FP6 programme (eRep project CIT5-028575, OpenKnowledge project FP6-027253 and SocialRep project MERG-CT-2005/029161). Jordi Sabater-Mir enjoys a RAMON Y CAJAL contract from the Spanish Government.

References

[1] J. Sabater, M. Paolucci, R. Conte, Repage: Reputation and image among limited autonomous partners, Journal of Artificial Societies and Social Simulation 9 (2) (2006). <http://jasss.soc.surrey.ac.uk/9/2/3.html>.

[2] S. Ramchurn, C. Sierra, L. Godo, N.R. Jennings, Devising a trust model for multi-agent interactions using confidence and reputation, International Journal of Applied Artificial Intelligence (18) (2004) 833–852.

[3] J. Sabater, C. Sierra, Review on computational trust and reputation models, Artificial Intelligence Review 24 (1) (2005) 33–60.

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013

[4] C. Dellarocas, The digitization of word-of-mouth: promise and challenges of online feedback, Working paper 4296-03, Massachusetts Institute of Technology (MIT), Sloan School of Management, 2004.
[5] J. Sabater, C. Sierra, Reputation and social network analysis in multi-agent systems, in: Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-02), Bologna, Italy, 2002, pp. 475–482.
[6] M. Schillo, P. Funk, M. Rovatsos, Who can you trust: dealing with deception, in: Proceedings of the Second Workshop on Deception, Fraud and Trust in Agent Societies, Seattle, USA, 1999, pp. 95–106.
[7] A. Abdul-Rahman, S. Hailes, Supporting trust in virtual communities, in: Proceedings of the Hawaii's International Conference on Systems Sciences, Maui, Hawaii, 2000.
[8] J. Carbo, J. Molina, J. Davila, Trust management through fuzzy reputation, International Journal in Cooperative Information Systems 1 (12) (2003) 135–155.
[9] D. Huynh, N.R. Jennings, N.R. Shadbolt, Developing an integrated trust and reputation model for open multi-agent systems, in: Proceedings of the Workshop on Trust in Agent Societies, Third International Joint Conference on Autonomous Agents and Multi Agent Systems, New York, USA, 2004, pp. 65–74.
[10] M. Miceli, C. Castelfranchi, The role of evaluation in cognition and social interaction, in: K. Dautenhahn (Ed.), Human Cognition and Agent Technology, Benjamins, Amsterdam, 2000.
[11] M. Grabisch, S.A. Orlovski, R.R. Yager, Fuzzy aggregation of numerical preferences, in: Fuzzy Sets in Decision Analysis, Operations Research and Statistics, 1998, pp. 31–68.
[12] eBay, eBay, 2006. <http://www.eBay.com>.
[13] Amazon, Amazon Auctions, 2006. <http://auctions.amazon.com>.
[14] BizRate, BizRate, 2006. <http://www.bizrate.com>.
[15] S. Sen, N. Sajja, Robustness of reputation-based trust: Boolean case, in: Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-02), Bologna, Italy, 2002, pp. 288–293.
[16] L. Mui, M. Mohtashemi, A. Halberstadt, A computational model for trust and reputation, in: Proceedings of the 35th Hawaii International Conference on System Sciences, 2002.
[17] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, 1988.
[18] J. Sabater, C. Sierra, Regret: a reputation model for gregarious societies, in: Proceedings of the Fourth Workshop on Deception, Fraud and Trust in Agent Societies, Montreal, Canada, 2001, pp. 61–69.
[19] J. Sabater, Trust and reputation for agent societies, Ph.D. thesis, Universitat Autonoma de Barcelona (UAB), 2003.
[20] M. Miceli, C. Castelfranchi, The role of evaluation in cognition and social interaction, in: Human Cognition and Agent Technology, Benjamin, Amsterdam, 2000.
[21] S. Philippe, Probability, possibility, belief: which and where? in: P. Smets (Ed.), Handbook of Defeasible Reasoning and Uncertainty Management Systems, Kluwer, Doordrecht, 1998, pp. 1–24.
[22] R. Yager, On the determination of strength of belief for decision support under uncertainity – Part I: Generating strength of belief, Fuzzy Sets and Systems 1 (2004) 117–128.
[23] R. Yager, Uncertainity modeling and decision support, Reliability Engineering and Systems Safety 85 (2004) 341–354.
[24] R. Yager, On the determination of strength of belief for decision support under uncertainity – Part II: Fusing strengths of belief, Fuzzy Sets and Systems 1 (2004) 129–142.
[25] V. Torra, Y. Narukawa, Modeling Decisions: Information Fusion and Aggregation Operators, Springer, 2007.

Please cite this article in press as: J. Sabater-Mir, M. Paolucci, On representation and aggregation ..., Int. J. Approx. Reason. (2007), doi:10.1016/j.ijar.2006.12.013