Information-Based Reputation

Carles Sierra[1] and John Debenham[2]

[1] Institut d'Investigacio en Intel.ligencia Artificial, Spanish Scientific Research Council, UAB
08193 Bellaterra, Catalonia, Spain sierra@iiia.csic.es
[2] University of Technology, Sydney, Australia debenham@it.uts.edu.au

Abstract. Information-based agents use tools from information theory to evaluate their utterances and to build their world model. When embedded in a social network these agents measure the strength of information flow in this sense. This leads to a model of information-based reputation in which agents share opinions, and observe the way in which their opinions effect the opinions of others. A method is proposed that supports the deliberative process of combining opinions into a group's reputation. The reliability of agents as opinion givers are measured in terms of the extent to which their opinions differ from that of the group reputation. These reliability measures are used to form an a priori reputation estimate given the individual opinions of a set of independent agents.

1 Introduction

Reputation measures are becoming a cornerstone of many applications over the web [1]. This is the case in recommender systems or in trading mediation sites. In these applications there is a need to assess, for instance, how much should we trust the recommendation coming from an unknown source, or how reliable a trading partner is. When the probability of having had previous interactions between two autonomous entities (agents), human or software, is very low, the use of reputation measures, i.e. group opinions, becomes a natural solution. In this paper we propose a reputation model that is inspired by information theory and that is based on the *information-based agency* explained elsewhere [2]. It also uses semantic distance over a shared ontology as a way to profit from similar experiences in the assessment of reputation, and we start to explore some social network analysis techniques to weigh the opinions of other agents according to their social relationships.

Reputation is the opinion (more technically, a social evaluation) of a group about something. So a group's reputation about a thing will be related in some way to the opinions that the individual group members hold towards that thing. An opinion is an assessment, judgement or evaluation of something, and are represented in this paper as probability distributions on a suitable ontology called the *evaluation space* E.

An opinion is an evaluation of an *aspect* of a thing. A rainy day may be evaluated as being "bad" from the aspect of being suitable for a picnic, and "good" from the aspect of watering the plants in the garden. An aspect is the "point of view" that an agent has when forming his opinion.

An opinion is evaluated in context. The *context* is the set of all things that the thing is being, explicitly or implicitly, evaluated with or against. The set of valuations of

all things in the context calibrates the valuation space. For example, "this is the best paper in the conference". The context can be vague: "of all the presents you could have given me, this is the best". If agents are to discuss opinions then they must have some understanding of each other's context.

Summarising the above, an *opinion* is an agent's evaluation of a particular aspect of a thing in context. A representation of an opinion will contain: the thing, its aspect, its context, and a distribution on E representing the evaluation of the thing.

In this paper we explore the case of opinions being formed through a social evaluation process illustrated in Figure 1. Each agent in a group of agents first forms an individual opinion on some thing. Second these individual opinions are shared with rest of the group. A group discussion follows as a result of which each agent states a revised opinion. Following that there is another discussion during which the group attempts to formulate a shared reputation for the thing. The model that we describe is based on three observations only for each participating agent: their initial individual opinion, their revised opinion, and the group's reputation if one is agreed upon. This social evaluation process was suggested by a process used to evaluate submissions to conferences.

2 The multiagent system

We assume that a multiagent system $\{\alpha, \beta_1, \ldots, \beta_o, \xi, \theta_1, \ldots, \theta_t\}$, contains an agent α that interacts with negotiating agents, β_i, information providing agents, θ_j, and an *institutional agent*, ξ, that represents the institution where we assume the interactions happen [3]. Institutions give a normative context to interactions that simplify matters (e.g an agent can't make an offer, have it accepted, and then renege on it). The institutional agent ξ may form opinions on the actors and activities in the institution and may publish reputation estimates on behalf of the institution. The agent ξ also fulfils a vital role to compensate for any lack of sensory ability in the other agents by promptly and accurately reporting observations as events occur; for an example, without such reporting an agent may have no way of knowing whether it is a fine day or not. When we consider the system from the point of view of a particular agent we will use agent α, and that is α's only significance.

Our agents are information-based [4], everything in their world is uncertain. To deal with this uncertainty, the world model, $\mathcal{M}^t$, consists of random variables each representing a point of interest in the world. Distributions are then derived for these variables on the basis of information received. Additionally, information-based agents [4] are endowed with machinery for valuing the information that they have, and that they receive. They were inspired by the observation that "everything an agent says gives away information". They model how much they know about other agents, and how much they believe other agents know about them. By classifying private information into functional classes, and by drawing on the structure of the ontology, they develop a map of the 'intimacy' [5] of their relationships with other agents.

2.1 Communication Model

We assume that all agents share an ontology O, that for simplicity we will consider as a set of well-formed expressions representing a given domain of discourse.[3]

An ontology is a tuple $O = (C, R, \leq, \sigma)$ where:

1. C is a finite set of concept symbols (including basic data types);
2. R is a finite set of relation symbols;
3. $\leq$ is a reflexive, transitive and anti-symmetric relation on C (a partial order)
4. $\sigma : R \rightarrow C^+$ is the function assigning to each relation symbol its arity

where $\leq$ is a traditional *is-a* hierarchy, and R contains relations between the concepts in the hierarchy.

Based on this ontology we define a simple language $\mathcal{C}$ that accounts for the expressions exchanged in gossiping dialogues, and is based on two fundamental primitives: experience$(\alpha, \beta, \varphi, \varphi')$ to represent, in φ, the world that β committed at bringing about and in φ' what α actually observed, and opinion$(\alpha, \beta, \varphi, o)$ to represent an opinion o that α makes about the behaviour or position of β with respect to φ. The opinion is expressed as a probability distribution p_i over a set of qualitative terms e_j. Experiences can also be considered argumentative moves in support of a particular opinion. Language $\mathcal{C}$ is then the set of utterances u defined as:

$$
\begin{aligned}
u &::= \text{inform}(agent, agent, content, time) \\
content &::= \text{opinion}(agent, agent, [term,](eval)) \mid \text{experience}(agent, agent, term, term) \\
term &::= \varphi | \phi | \ldots (*\textit{expression from ontology } O*) \\
eval &::= e = p \mid e = p, eval \\
e &::= good \mid bad \mid \ldots (*\textit{qualitative term}*) \\
p &::= \textit{a point in } [0,1] \\
time &::= \textit{a point in time} \\
agent &::= \alpha \mid \beta \mid \ldots (*\textit{agent identifiers}*)
\end{aligned}
$$

We will note by $\mathcal{A}^t$ the set of existing agents at instant t, by $\mathcal{E}$ the set of all possible evaluation values e, and by Φ the set of all ontology-complaint terms. For example:

$\text{inform}(John, me, \text{opinion}(John, Carles, wrapping(package), (ghastly = 0.7)), t)$
$\text{inform}(John, me, \text{opinion}(Carles, John, suggesting(wine(Margaret\ River)), (excellent = 0.9)), t)$
$\text{inform}(John, me, \text{experience}(John, Carles, package(date(Monday)), package(date(Friday)), t)$
$\text{inform}(John, me, \text{experience}(John, Carles, fly(elephant), \neg fly(elephant)), t)$

The concepts within an ontology are closer, semantically speaking, depending on how far away are they in the structure defined by the $\leq$ relation. Semantic distance plays

[3] Local ontologies could also be considered together with appropriate ontology alignment techniques [6].

a fundamental role in strategies for information-based agency. How stated opinions, opinion$(\cdot)$, about objects in a particular semantic region, and their subsequent validation *affect* our decision making process about the significance of future opinions on nearby semantic regions is crucial to model the common sense that human beings apply. A measure [7] bases the *semantic similarity* between two concepts on the path length induced by $\leq$ (more distance in the $\leq$ graph means less semantic similarity), and the *depth* of the subsumer concept (common ancestor) in the shortest path between the two concepts (the deeper in the hierarchy, the closer the meaning of the concepts). Semantic similarity could then be defined as:

$$\mathrm{Sim}(c,c') = e^{-\kappa_1 l} \cdot \frac{e^{\kappa_2 h} - e^{-\kappa_2 h}}{e^{\kappa_2 h} + e^{-\kappa_2 h}}$$

where l is the length (i.e. number of hops) of the shortest path between the concepts, h is the depth of the deepest concept subsuming both concepts, and κ_1 and κ_2 are parameters scaling the contribution of shortest path length and depth respectively.

The following does not depend on this particular definition. Suppose an ontology is populated with probability distributions at each branch representing the preference in some sense of an agent; e.g. if wine $\geq$ {red wine, white wine} then the probability at that branch could represent Carles' preference for red or white wine. If the same ontology is populate with John's probabilities then a metric such as the Kullback-Leibler [8] divergence can be used to measure the difference in the significance of the term 'red wine' to Carles and to John.

2.2 The Social Structure of the Multiagent System

Agents, or groups, in an evolving network can be described by a number of measures of their importance or prominence [9] [10]. These measures summarise the structural relations among all nodes in the network and account for an agent's choices (whom do I link to) as well as the other agent's choices (who links to me). Centrality measures try to determine prominence by not taking into account the direction of the ties, and prestige measures when direction matters. Given a matrix $R(n,n)$ that represents in $r_{ij} \in [0,1]$ the intensity of the relation R from i to j we define:

- Centrality measures. Determining in how many relationships a particular agent is involved.
  - Normalised Degree Centrality. The extent to which a node connects to the rest.

$$C_d(i) = \frac{\sum_{j=1}^{n} r_{ij}}{n-1}$$

  - Normalised Closeness Centrality. How near a node is from the rest.

$$C_c(i) = \frac{n-1}{\sum_{j=1}^{n} d(i,j)}$$

    where $d(i,j)$ is the minimum distance between i and j in the graph

- Normalised Betweenness Centrality. The extent to which an agent lies on the shortest paths between pairs of agents in the graph.

$$C_b(i) = \frac{2}{(n-1)(n-2)} \cdot \sum_{j,k \neq i, j \neq k} \frac{s_{jk}(i)}{s_{jk}}$$

  where $s_{jk}(i)$ is the number of shortest paths between j and k including i, and s_{jk} is the total number of shortest paths between j and k.

– Prestige Degree. Determining how many links an agent receives.

$$P(i) = \frac{\sum_{j=1}^{n} r_{ji}}{n-1}$$

The preceding measures are topological and do not capture what the connections between the individual agents are used for. From the perspective of information-based agency we are interested in two things: first, how much information is passing along the connections, and second, the value of that information to the receiving agent.

The 'Source Coding Theorem' of Shannon states that N independent, identically-distributed random variable each with entropy $\mathbb{H}(X)$ can be compressed into marginally more than $N \times \mathbb{H}(X)$ bits of information. In other words, if we know the amount of information that has been transmitted in bits, and that the coding is loss-less then we know the amount of information that has been transmitted in terms of the lack of uncertainty that it could bring. Further, if we have mutually exclusive events, E_i, each with prior probabilities, p_i, then the *expected information content* $\mathbb{I}$ of a message that transforms the priors p_i into posterior probabilities q_i is: $\mathbb{I} = \sum_i q_i \times \log(\frac{q_i}{p_i})$. These ideas enable use to analyse network structure from the perspective of information flow. [11] defines the *path-transfer centrality* of vertex i as $-\sum_j p_{ij} \log p_{ij}$ where p_{ij} is the probability that a communication path starting at node i will end at node j. If an agent receives a message containing information $\mathbb{I}$ then the Shannon *value* of $\mathbb{I}$ is: $\mathbb{H}(\mathcal{M}^t \,|\, I) - \mathbb{H}(\mathcal{M}^t)$, where $\mathcal{M}^t$ is the agent's world model. When used together with the ontology and a map of $\mathcal{M}^t$ that categorises the agent's information, this measure can be used to take stock of the information in $\mathcal{M}^t$.

3 Forming Opinions

This section describes how an information-based agent forms opinions. Section 4 will describe how the opinions of the agents in a group may be distilled into a reputation.

An opinion is a valuation by an agent of an aspect of a thing taken in context. Formally, $O_i(z,a,C)$ represents the result of the valuation by agent β_i of aspect a of thing z in context C. For example, the valuation by agent "Carles" of the "scientific quality" aspect of the thing "John's paper" in the context of "the AAMAS conference submissions". Opinions are communicated using the language described in Section 2.1. The context C is often subjectively chosen by the agent, and is not part of the opinion$(\cdot)$ primitive, although context may be the subject of associated argumentation. For example, re-using an example of communication from Section 2.1:

$$\text{inform}(John, me, \text{opinion}(John, Carles, wrapping(package), (ghastly = 0.7)), t)$$

we can extract an opinion as:

$$O(package, wrapping, the\ way\ I\ do\ wrapping) = (ghastly = 0.7)$$

As noted above, to preserve consistency and generality we assume that all opinions are expressed as probability distributions over some suitable E. If an agent expresses an opinion as $\mathbb{P}(X = e_i)$ we treat this as the distribution with minimum relative entropy with respect to the prior subject to the constraint $\mathbb{P}(X = e_i)$ — in case there is no known prior we use the maximum entropy, uniform distribution. For example, if $E = (\text{fine}, \text{cloudy}, \text{wet}, \text{storm})$ then the opinion "I am 70% certain that tomorrow will be fine" will be represented as $(0.7, 0.1, 0.1, 0.1)$ for a uniform prior.

The distributions in an agent's world model $\mathcal{M}^t$ represent the agent's opinions about the value of the corresponding random variable over some valuation space. Opinions may be derived from opinions. For example, to form an opinion on "tomorrow's suitability for a picnic" and agent may introduce random variables for: tomorrow's mid-day temperature, tomorrow's mid-day cloud cover, and tomorrow's mid-day wind strength, construct distributions for them using on-the-fly weather forecast information, and then derive an opinion about the picnic somehow from these three distributions.

In Section 3.1 we describe how the distributions in the world model are updated as real-time information becomes available; in that section we also estimate the reliability of each information source by subsequently validating the information received from it.

3.1 Updating Opinions with Real-Time Information

In the absence of in-coming messages the distributions in $\mathcal{M}^t$ should gradually decay towards some zero-information state. In many cases there is background knowledge about the world — for example, a distribution of the daily maximum temperature in Barcelona in May — such a distribution is called a *decay-limit distribution*. If the background knowledge is incomplete then one possibility is to assume that the decay limit distribution has maximum entropy whilst being consistent with the available data. Given a distribution, $\mathbb{P}(X_i)$, and a decay limit distribution $\mathbb{D}(X_i)$, $\mathbb{P}(X_i)$ decays by:

$$\mathbb{P}^{t+1}(X_i) = \Delta_i(\mathbb{D}(X_i), \mathbb{P}^t(X_i)) \tag{1}$$

where Δ_i is the *decay function* for the X_i satisfying the property that $\lim_{t\to\infty} \mathbb{P}^t(X_i) = \mathbb{D}(X_i)$. For example, Δ_i could be linear: $\mathbb{P}^{t+1}(X_i) = (1-\nu_i) \times \mathbb{D}(X_i) + \nu_i \times \mathbb{P}^t(X_i)$, where $\nu_i < 1$ is the decay rate for the i'th distribution. Either the decay function or the decay limit distribution could also be a function of time: Δ_i^t and $\mathbb{D}^t(X_i)$.

The following procedure updates $\mathcal{M}^t$. Suppose that α receives a message μ from agent β at time t.[4] Suppose that this message states that something is so with probability v, and suppose that α attaches an epistemic belief $\mathbb{R}^t(\alpha, \beta, \mu)$ to μ — this probability reflects α's level of personal *caution*. Each of α's active plans, s, contains constructors for a set of distributions $\{X_i\} \in \mathcal{M}^t$ together with associated *update functions*, $J_s(\cdot)$,

[4] This message is not necessarily a message from the language in section 2.1. We refer with μ to any *inform* message with propositional content that can be processed by the agent.

such that $J_s^{X_i}(\mu)$ is a set of linear constraints on the posterior distribution for X_i. Denote the prior distribution $\mathbb{P}^t(X_i)$ by p, and let $p_{(\mu)}$ be the distribution with minimum relative entropy[5] with respect to p: $p_{(\mu)} = \arg\min_r \sum_j r_j \log \frac{r_j}{p_j}$ that satisfies the constraints $J_s^{X_i}(\mu)$. Then let $q_{(\mu)}$ be the distribution:

$$q_{(\mu)} = \mathbb{R}^t(\alpha,\beta,\mu) \times p_{(\mu)} + (1 - \mathbb{R}^t(\alpha,\beta,\mu)) \times p \tag{2}$$

and then let:

$$\mathbb{P}^t(X_{i(\mu)}) = \begin{cases} q_{(\mu)} & \text{if } q_{(\mu)} \text{ is more interesting than } p \\ p & \text{otherwise} \end{cases} \tag{3}$$

A general measure of whether $q_{(\mu)}$ is more interesting than p is: $\mathbb{K}(q_{(\mu)} \| \mathbb{D}(X_i)) > \mathbb{K}(p \| \mathbb{D}(X_i))$, where $\mathbb{K}(x\|y) = \sum_j x_j \ln \frac{x_j}{y_j}$ is the Kullback-Leibler divergence between two probability distributions x and y.

Finally merging Eqn. 3 and Eqn. 1 we obtain the method for updating a distribution X_i on receipt of a message μ:

$$\mathbb{P}^{t+1}(X_i) = \Delta_i(\mathbb{D}(X_i), \mathbb{P}^t(X_{i(\mu)})) \tag{4}$$

This procedure deals with integrity decay, and with two probabilities: first, the probability v in the message μ, and second the belief $\mathbb{R}^t(\alpha,\beta,\mu)$ that α attached to μ.

Reliability of the Information Source. An empirical estimate of $\mathbb{R}^t(\alpha,\beta,\mu)$ may be obtained by measuring the 'difference' between commitment and verification. Suppose that μ is received from agent β at time u and is verified by ξ as μ' at some later time t.[6] Denote the prior $\mathbb{P}^u(X_i)$ by p. Let $p_{(\mu)}$ be the posterior minimum relative entropy distribution subject to the constraints $J_s^{X_i}(\mu)$, and let $p_{(\mu')}$ be that distribution subject to $J_s^{X_i}(\mu')$. We now estimate what $\mathbb{R}^u(\alpha,\beta,\mu)$ should have been in the light of knowing *now*, at time t, that μ should have been μ'.

The idea of Eqn. 2, is that $\mathbb{R}^t(\alpha,\beta,\mu)$ should be such that, *on average* across $\mathcal{M}^t$, $q_{(\mu)}$ will predict $p_{(\mu')}$ — no matter whether or not μ was used to update the distribution for X_i, as determined by the condition in Eqn. 3 at time u. The *observed reliability* for μ and distribution X_i, $\mathbb{R}^t_{X_i}(\alpha,\beta,\mu)|\mu'$, on the basis of the verification of μ with μ', is the value of k that minimises the Kullback-Leibler divergence:

$$\mathbb{R}^t_{X_i}(\alpha,\beta,\mu)|\mu' = \arg\min_k \mathbb{K}(k \cdot p_{(\mu)} + (1-k) \cdot p \parallel p_{(\mu')})$$

[5] Given a probability distribution q, the *minimum relative entropy distribution* $p = (p_1, \ldots, p_I)$ subject to a set of J linear constraints $g = \{g_j(p) = a_j \cdot p - c_j = 0\}, j = 1, \ldots, J$ (that must include the constraint $\sum_i p_i - 1 = 0$) is: $p = \arg\min_r \sum_j r_j \log \frac{r_j}{q_j}$. This may be calculated by introducing Lagrange multipliers λ: $L(p,\lambda) = \sum_j p_j \log \frac{p_j}{q_j} + \lambda \cdot g$. Minimising L, $\{\frac{\partial L}{\partial \lambda_j} = g_j(p) = 0\}, j = 1, \ldots, J$ is the set of given constraints g, and a solution to $\frac{\partial L}{\partial p_i} = 0, i = 1, \ldots, I$ leads eventually to p. Entropy-based inference is a form of Bayesian inference that is convenient when the data is sparse [12] and encapsulates common-sense reasoning [13].

[6] This could be later communicated as inform$(\gamma, \alpha, \text{experience}(\gamma,\beta,\mu,\mu'), t)$.

The predicted *information* in the enactment of μ with respect to X_i is:

$$\mathbb{I}^t_{X_i}(\alpha,\beta,\mu) = \mathbb{H}^t(X_i) - \mathbb{H}^t(X_{i(\mu)}) \tag{5}$$

that is the reduction in uncertainty in X_i where $\mathbb{H}(\cdot)$ is Shannon entropy. Eqn. 5 takes account of the value of $\mathbb{R}^t(\alpha,\beta,\mu)$.

If $\mathbf{X}(\mu)$ is the set of distributions that μ affects, then the *observed reliability* of β on the basis of the verification of μ with μ' is:

$$\mathbb{R}^t(\alpha,\beta,\mu)|\mu' = \frac{1}{|\mathbf{X}(\mu)|}\sum_i \mathbb{R}^t_{X_i}(\alpha,\beta,\mu)|\mu' \tag{6}$$

If $\mathbf{X}(\mu)$ are independent the predicted *information* in μ is:

$$\mathbb{I}^t(\alpha,\beta,\mu) = \sum_{X_i \in \mathbf{X}(\mu)} \mathbb{I}^t_{X_i}(\alpha,\beta,\mu) \tag{7}$$

Suppose α sends message μ to β where μ is α's private information, then assuming that β's reasoning apparatus mirrors α's, α can estimate $\mathbb{I}^t(\beta,\alpha,\mu)$.

For each formula φ at time t when μ has been verified with μ', the *observed reliability* that α has for agent β in φ is:

$$\mathbb{R}^{t+1}(\alpha,\beta,\varphi) = (1-\nu)\times\mathbb{R}^t(\alpha,\beta,\varphi) + \nu\times\mathbb{R}^t(\alpha,\beta,\mu)|\mu'\times \mathrm{Sim}(\varphi,\mu)$$

where Sim measures the semantic distance between two sections of the ontology as introduced in Section 2.1, and ν is the learning rate. Over time, α notes the context of the various μ received from β, and over the various contexts calculates the relative frequency, $\mathbb{P}^t(\mu)$. This leads to an overall expectation of the *reliability* that agent α has for agent β:

$$\mathbb{R}^t(\alpha,\beta) = \sum_\mu \mathbb{P}^t(\mu)\times\mathbb{R}^t(\alpha,\beta,\mu)$$

3.2 Verifiable Opinions

An opinion is *verifiable* if within a "reasonable amount of time" it ceases to be an opinion and becomes an observable fact; for example, the opinion "tomorrow's maximum temperature will be over 30°" is verifiable, whereas the opinion "the Earth will exist in 100,000 years time" is not verifiable in any practical sense, and "Brahms' symphonies are ghastly" will never be verifiable.

The articulation by β of a verifiable opinion carries with it the intrinsic commitment that it will in due time become an observable true fact. α will be interested in any variation between β's commitment, φ, and what is actually observed (as advised by the institution agent ξ), as the fact, φ'. We denote the relationship between opinion and fact, $\mathbb{P}^t(\mathrm{Observe}(\varphi')|\mathrm{Commit}(\varphi))$ simply as $\mathbb{P}^t(\varphi'|\varphi) \in \mathcal{M}^t$.

In the absence of in-coming messages the conditional probabilities, $\mathbb{P}^t(\varphi'|\varphi)$, should tend to ignorance as represented by the *decay limit distribution* and Eqn. 1. We now show how Eqn. 4 may be used to revise $\mathbb{P}^t(\varphi'|\varphi)$ as observations are made. Let the

set of possible factual outcomes be $\Phi = \{\varphi_1, \varphi_2, \ldots, \varphi_m\}$ with prior distribution $p = \mathbb{P}^t(\varphi'|\varphi)$. Suppose that message μ is received from ξ that verifies or refutes a previously stated verifiable opinion expressed by β, we estimate the posterior $p_{(\mu)} = (p_{(\mu)i})_{i=1}^m = \mathbb{P}^{t+1}(\varphi'|\varphi)$.

First, if $\mu = (\varphi_k, \varphi)$ is observed then α may use this observation to estimate $p_{(\varphi_k)k}$ as some value d at time $t+1$. We estimate the distribution $p_{(\varphi_k)}$ by applying the principle of minimum relative entropy as in Eqn. 4 with prior p, and the posterior $p_{(\varphi_k)} = (p_{(\varphi_k)j})_{j=1}^m$ satisfying the single constraint: $J^{(\varphi'|\varphi)}(\varphi_k) = \{p_{(\varphi_k)k} = d\}$.

Second, we consider the effect that the verification ϕ' of another simple, verifiable opinion ϕ of β has on p. This is achieved by appealing to the structure of the ontology using the $\mathrm{Sim}(\cdot)$ function. Given the observation $\mu = (\phi', \phi)$, define the vector t by:

$$t_i = \mathbb{P}^t(\varphi_i|\varphi) + (1 - \mid \mathrm{Sim}(\phi', \phi) - \mathrm{Sim}(\varphi_i, \varphi) \mid) \cdot \mathrm{Sim}(\varphi', \phi)$$

for $i = 1, \ldots, m$. t is not a probability distribution. The multiplying factor $\mathrm{Sim}(\varphi', \phi)$ limits the variation of probability to those formulae whose ontological context is not too far away from the observation. The posterior $p_{(\phi', \phi)}$ is defined to be the normalisation of t.

In this section we have shown how an information-based agent models the accuracy of an agent's opinions when they are verifiable. The model produced is predictive in the sense that when an opinion is stated it gives a distribution of expectation over the space of factual outcomes.

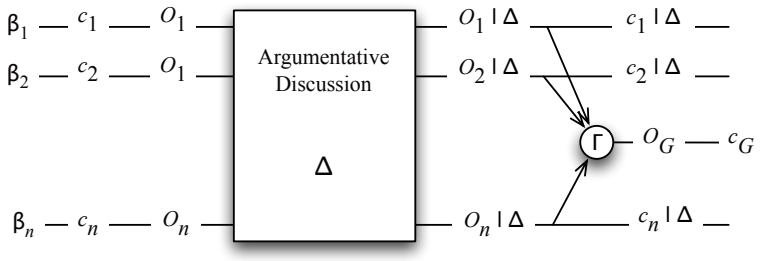
3.3 Unverifiable Opinions

If an opinion can not be verified then one way in which it may be evaluated is to compare it with the corresponding individual opinions, or group reputation, of a group of agents. The focus of this paper is on reputation; that is, a social evaluation conducted by a group. We deal with unverifiable opinions using a social evaluation framework that is abstracted from any particular case and is illustrated in Figure 1. The idea is that a group G of n agents independently form a prior opinion, O_i on the same thing. Each agent has a prior confidence value, c_i, that estimates how close its prior opinion, O_i, is expected to be to the reputation, or common opinion, of the group, R_G — precisely c_i measures how effective the agent is at influencing the opinions of other agents, it does *not* measure how good its opinion is in any absolute sense as the opinion is assumed to be unverifiable. The agents then make their prior opinions public to the other agents and an argumentative discussion, Δ, takes place during which the agents may choose to revise their opinions, $O_i|\Delta$. When the revised opinions are published a second argumentative discussion, Γ, takes place during which the agents attempt to distil their opinions into a group reputation, R_G. The confidence estimates, c_i are then revised by noting the differences between O_i, $O_i|\Delta$ and R_G, to give posterior values, $c_i|\Delta$. The processes in Figure 1 are summarised as:

$$\Delta : f(\{(O_i, c_i\}) = \{O_i|\Delta\}$$

$$\Gamma : g(\{(O_i|\Delta, c_i\}) = (R_G, d_G)$$

Fig. 1. The social evaluation framework in which a group G of n agents $\beta_1,\ldots,\beta_n$ table their private opinions $O_1,\ldots,O_n$, have an open, argumentative discussion Δ (see Section 3.3), and then revise their opinions $O_1|\Delta,\ldots,O_n|\Delta$. This is followed by another argumentative discussion Γ (see Section 4) during which the agents consider whether revised opinions can be distilled into a common reputation R_G. The symbols c_i and c_G are confidence values as explained below.



$$\{\Delta,\Gamma\} : h(\{(O_i,c_i,O_i|\Delta\},R_G) = \{c_i|\Delta\}$$

The function $f(\cdot)$ is the product of the discussion Δ — we simply observe the outcome. Function $g(\cdot)$ is described in Section 4, and $h(\cdot)$ in Section 5.

4 Combining Opinions and Forming Reputation

A reputation is a social evaluation by a group. When the group is a set of autonomous agents the only sense in which an opinion can exist is as a common opinion throughout the group. The objective of the argumentative process Γ in Figure 1 is to determine a common view if one exists. The following procedure first determines whether a common view exists, and second it offers three views of what that common view could be. The three different views vary with differing degrees of statistical dependence between the agents.

The process of distilling opinions into a reputation can not simply be computed. For example, consider two agents who are reviewing the same conference paper and are in total agreement about the result "a 'strong accept' with confidence 0.8" where the reliability of each agent is 90%. What should their combined opinion, or in this case 'paper reputation', be? As their individual reliability is 90% perhaps the common view is "a 'strong accept' with confidence 0.72". Alternatively because they both agree, and may have quite different reasons supporting their views, perhaps the common view should be "a 'strong accept' with some confidence greater than 0.8".

The work described in the remainder of this section and in Section 5 is expressed in terms of two agents; it extends naturally to n agents. The procedure is based on three methods that are detailed below.

Dependent Method. To form a *combined opinion* of two opinions, X_1 and X_2, construct the joint distribution $W = (X_1, X_2, Z)$ and impose the constraints:

$$\left(\sum_i \mathbb{P}(W = w_i) \mid X_k = x_j\right) = \mathbb{P}(X_k = x_j),\ k = 1,2$$

$$\left(\sum_i \mathbb{P}(W = w_i) \mid X_k = Z\right) = c_k,\ k = 1,2$$

let W be the distribution of maximum entropy that satisfies these constraints. Then the combined opinion $\mathrm{Dep}(X_1, X_2)$ is $\mathbb{P}(Z = z)$. If the data is inconsistent then the value is undefined — this is a test of whether the data is consistent. If the data is inconsistent then this indicates that there is no shared opinion. Being based on a maximum entropy calculation the posterior is a conservative combination of the given opinions — it is "maximally noncommittal" to that which is not known. To calculate this dependent, combined opinion when the prior is known, calculate the minimum relative entropy distribution with respect to that prior using the same constraints as described.

Υ *Method.* Let's define $\mathbb{P}(\alpha, d)$ as the probability that an opinion O_α expressed by α (i.e, a probability distribution) is at distance d of the true distribution (or at distance d of a group opinion). That is, the probability that a certain distribution Q is the right one is defined as $P(Q \text{ is right}) = \mathbb{P}(\alpha, DIST(O_\alpha, Q))$ for an appropriate distance measure $DIST$.[7] These distributions can be obtained by datamining past group opinion formation processes.

Given a group G, we look for the group opinion, R_G such that the certainty on that group opinion being the right one is maximised. That is,

$$R_G = \max_Q \Upsilon(\{\mathbb{P}(\alpha, DIST(O_\alpha, Q))\}_{\alpha \in G})$$

Where Υ is the uninorm operator [15]. In case there are several such group opinions we prefer the one with maximum entropy. And then,

$$d_G = \Upsilon(\{\mathbb{P}(\alpha, DIST(O_\alpha, R_G))\}_{\alpha \in G})$$

For the values in Table 1, we discreetise the $\mathbb{P}(\alpha, d)$ in the intervals between the points in the following list: $[0, 0.035, 0.3, 0.5, 0.8, 1]$.

Independent Method. Given a prior distribution $\mathbb{P}(W = x_j)$, a pair of opinions, $\mathbb{P}(X_i = x_j)\ i = 1,2$, with their respective certainties c_i, assuming that the agents' opinions are statistically independent, let $w_{i,j} = c_i \times \mathbb{P}(X_i = x_j),\ i = 1,2$, and let $v_j = \frac{\prod_i w_{i,j}}{\prod_i w_{i,j} + \prod_i (1 - w_{i,j})}$ then the combined opinion $\mathrm{Ind}(X_1, X_2)$ is: $v_j + (1 - \sum_k v_k) \times \mathbb{P}(W = x_j)$, with strength $\sum_k v_k$. This method assumes that the priors are independent (unlikely in practice) and has the property that the probabilities in two similar distributions are amplified.

The overall procedure plays the role of a mediator. If the 'Dependent Method' does not return a value then the data is inconsistent, and the agents should either have further

[7] Kullback-Leibler divergence, or the earth movers distance [14] could be used.

discussion or "agree to disagree". Otherwise calculate the three values Dep$(\cdot)$, $\Upsilon(\cdot)$ and Ind$(\cdot)$. Propose $\Upsilon(\cdot)$ to the agents, and if they accept it then that is their common opinion. Otherwise propose that their common opinion lies somewhere between Dep$(\cdot)$ and Ind$(\cdot)$ and leave it to them to determine it.

Table 1 contains some sample values for the three methods. In Case 3 the two opinions are identical with maximal value of 0.8 and strengths of 0.8 and 0.9. The Dep(X_1, X_2) method is conservative and gives 0.77 because of the strength values. The $\Upsilon(X_1, X_2)$ method balances the strength uncertainty with the fact that their are two shared views to give 0.8. The Ind(X_1, X_2) method is bold and gives 0.85 because two agents share the same view; the boldness of the Ind(X_1, X_2) method is balanced by its comparatively low strength values.

Table 1. Three cases of sample values for the three methods for combining opinions. In each case the opinions are X_1 and X_2 and the strength of the distributions is denoted by "Str". The right hand column contains the discreetised $\mathbb{P}(\alpha, d)$ values described in the 'Υ Method'. All calculations were performed with a uniform prior.

Case 1	X_1	0.1000	0.5000	0.2000	0.1000	0.1000	Str = 0.9	$P = \langle 0.9, 0.05, 0.03, 0.01, 0.01 \rangle$
	X_2	0.0500	0.8000	0.0500	0.0500	0.0500	Str = 0.7	$P = \langle 0.7, 0.2, 0.05, 0.03, 0.02 \rangle$
	Dep	0.0919	0.5590	0.1653	0.0919	0.0919	$c_G \approx 1$	
	Υ	0.0700	0.7000	0.1700	0.0700	0.0700	$c_G = 0.95$	
	Ind	0.0978	0.6044	0.1022	0.0978	0.0978	$c_G = 0.53$	
Case 2	X_1	0.1000	0.6000	0.1000	0.1000	0.1000	Str = 0.8	$P = \langle 0.8, 0.1, 0.04, 0.01, 0.01 \rangle$
	X_2	0.0500	0.8000	0.0500	0.0500	0.0500	Str = 0.9	$P = \langle 0.9, 0.06, 0.03, 0.01, 0.01 \rangle$
	Dep	0.0683	0.7266	0.0683	0.0683	0.0683	$c_G \approx 1$	
	Υ	0.08	0.63	0.08	0.08	0.08	$c_G = 0.97$	
	Ind	0.0601	0.7596	0.0601	0.0601	0.0601	$c_G = 0.72$	
Case 3	X_1	0.0500	0.8000	0.0500	0.0500	0.0500	Str = 0.8	$P = \langle 0.8, 0.1, 0.04, 0.01, 0.01 \rangle$
	X_2	0.0500	0.8000	0.0500	0.0500	0.0500	Str = 0.9	$P = \langle 0.9, 0.06, 0.03, 0.01, 0.01 \rangle$
	Dep	0.0573	0.7707	0.0573	0.0573	0.0573	$c_G \approx 1$	
	Υ	0.05	0.8	0.05	0.05	0.05	$c_G = 0.97$	
	Ind	0.0363	0.8548	0.0363	0.0363	0.0363	$c_G = 0.83$	

5 Reputation of the Agents

In the previous section we described how a mediator could assist agents to agree on a common opinion, or reputation, of some thing being evaluated. Additionally, the institution ξ builds a view of the reputation of the individual agents who perform the evaluations by observing the process illustrated in Figure 1. In particular, ξ observes the development of the c_i values (described below), the distances between initial opinion O_i and considered opinion $O_i|\Delta$, and the distances between both opinions and the group reputation R_G when it exists.

Given two opinions X_1 and X_2 the *strength* of X_1 on X_2 is defined as: $\mathbb{P}(X_1 = X_2)$. If X_1 and X_2 are both defined over the same valuation space $E = \{e_i\}_{i=1}^n$ then: $\mathbb{P}(X_1 =$

$X_2) = \sum_i P(W = w_i) \mid X_1 = X_2$, where $W = (X_1, X_2)$ is the joint distribution. That is, we sum along the diagonal of the joint distribution. We estimate the diagonal w_i values using the dependent estimate: $\mathbb{P}(X_1 = e_i) \wedge \mathbb{P}(X_1 = e_i) = \min_j \mathbb{P}(X_j = e_i)$, and hence: $\mathrm{Str}(X_1, X_2) = \mathbb{P}(X_1 = X_2) = \sum_i \min_j \mathbb{P}(X_j = e_i)$. A measure of the distance between X_1 and X_2 is then: $\mathrm{Dist}(X_1, X_2) = 1 - \mathrm{Str}(X_1, X_2)$. This definition of strength is consistent with the 'Dependent Method' in Section 4 that is the basis of the reputation mediation procedure. Other definitions include the Kullback-Leibler divergence, $\mathrm{Dist}(X_1, X_2) = \mathbb{K}(X_1 || X_2)$, and the earth movers distance [14].

Each time a reputation R_G is formed, the c_i values are updated using: $c_i|\Delta = \mu \times \mathrm{Dist}(O_i, R_G) + (1 - \mu) \times c_i$, where μ is the learning rate. These c_i values are the product of successive social evaluation processes, and so they are reputation estimates.

The measures described above do not take the structure of the evaluation space E into account. Four additional measures are:

A generic distance measure. $\mathrm{Dist}(X, Y) = \mathbb{K}(X' || Y')$ where (X', Y') is a permutation of (X, Y) the satisfies $X' < Y'$, and the order is defined by: $R_G < O_i|\Delta < O_i$. I.e. the earliest occurring distribution "goes in the second argument". This complication with ordering is necessary because $\mathbb{K}$ is not symmetric; it attempts to exploit the sense of relative entropy. An alternative is to use the symmetric form as it was originally proposed: $\frac{1}{2}(\mathbb{K}(X, Y) + \mathbb{K}(Y, X))$

A distance measure when the prior, Z, is known. This builds on the generic measure, and captures the idea that the distance between a pair of unexpected distributions is greater than the difference between a pair of similar, expected distributions. We measure of how expected X is by: $\mathbb{K}(X, Z)$, and normalise it by: $max_I \mathbb{K}(I, Z)$ to get: $e(X) = \frac{\mathbb{K}(X,Z)}{max_I \mathbb{K}(I,Z)}$. Then this measure is the arithmetic product of the previous generic measure with: $\frac{e(X)+e(Y)}{2}$.

A semantic distance measure. Suppose there is a difference measure $\mathrm{Diff}(\cdot, \cdot)$ defined between concepts in the ontology — it could be related to the $\mathrm{Sim}(\cdot, \cdot)$ function in Section 2.1. Then the distance between two opinions X and Y over valuation space E (represented as distributions p_i and q_i respectively) is: $\mathrm{Dist}(X, Y) = \sum_{ij} p_i \times q_j \times \mathrm{Diff}(e_i, e_j)$ where e_i are the categories in E.

A distance measure when E is ordered and the prior is known. If the valuation space E has a natural order, and if there is a known prior then define $\mathrm{Diff}(e_i, e_j)$ to be the proportion of the population that is expected to lie between e_i and e_j. Then define $\mathrm{Dist}(X, Y) = \sum_{ij} p_i \times q_j \times \mathrm{Diff}(e_i, e_j)$. For example, in conference reviewing, if the expectation is that 40% of reviews are 'weak accept' and 20% are 'accept' then Diff('weak accept', 'accept')$= \frac{40}{2} + \frac{20}{2}$; i.e. taking the mid points of the intervals.

The measures described for $\mathrm{Dist}(X, Y)$ are now used to enable ξ to attribute various reputations to agents. These reputation measures all assume that the agents have been involved in a number of successive social evaluation rounds as shown in Figure 1.

Inexorable. If agent β_i is such that: $\mathrm{Dist}(O_i, O_i|\Delta) \ll \mathrm{Dist}(O_i, O_j|\Delta), \forall j \neq i$ consistently holds then β_i is *inexorable*.

Predetermination. If: $\mathrm{Dist}(O_i, R_G) \ll \mathrm{Dist}(O_j, R_G), \forall j \neq i$ consistently, then β_i is a good '*predeterminer*'. Such an agent will have a high c_i value.

Persuasiveness. If β_i is such that: $\mathrm{Dist}(O_i, O_j|\Delta) \ll \mathrm{Dist}(O_j, O_j|\Delta), \forall j \neq i$ consistently then β_i is *persuasive*.

Compliance. If β_i is such that: $O_i|\Delta \approx \arg\min_X \sum_{j \neq i} \mathrm{Dist}(O_j|\Delta, X)$, then β_i is *compliant*.

Dogmatic. If β_i is such that: $O_i = O_i|\Delta$ consistently then β_i is *dogmatic*. A dogmatic agent is highly inexorable.

Adherence. If β_i is such that $O_i|\Delta = O_j$ where $j = \arg\max_{k,k \neq i} c_k$ consistently then β_i is *adherent* (in this round adherent to agent β_j).

6 Discussion

This paper has proposed a number of methods to ground the social building of reputation measures. The methods are based on information theory and permit to combine opinions when there is a high level of independence in the formation of the individual opinions. The method permits the computation of reputation values as aggregation of individual opinions, and also detects when agreement is not feasible. This impossibility may be used to trigger further discussions among the members of the group or to introduce changes in the composition of the group to permit agreements.

The use of social network analysis measures permits to define heuristics on how to combine opinions when there is no complete independence in the opinions expressed by the agents. There are a number of different relationships that may be used to guess dependency. For instance, in the context of scientific publications, co-authorship or affiliation, meaning that authors have written papers together or belong to the same laboratory may indicate a significant exchange of information between them and therefore a certain level of dependency. The aggregation of values by function h can then use these measures to diminish the joint influence of dependent opinions into the reputation. This is to be explored in future extensions of the information based reputation model.

Also, social networks can be used to assess initial values of c_i, the confidence on agent's opinions. For instance, we can say that an individual is expert in an area (keyword) if it is author of highly cited papers on the topic, has reviewed prestigious papers on the area, and has a central role in the college. This is easily expressed as

$$c_i = f\left(\sum_{\substack{(i,p)\in Authorship,\\(p,X)\in Area}} P^{Citation}(p), \sum_{\substack{(i,p)\in Review,\\(p,X)\in Area}} P^{Citation}(p), C_b^{College}(i)\right)$$

where $(i,p) \in Authorship$ means that agent i is author of paper p, $(p,X) \in Area$ means that paper p is on topic X and $(i,p) \in Review$ means that agent i has reviewed paper p. $Citation$ relates papers and $College$ relates authors. See Section 2.2 for definitions of P and C_b.

Our future work will include the in depth analysis of Social Network Measures in the information model reputation and the experimental analysis of the model in the context of scientific publishing as planned in the LiquidPub project (http://www.liquidpub.org).

Also, we will analyse the robustness of the proposed model in front of strategic reasoners that may try and manipulate the scores to their benefit.

References

1. Sabater, J., Sierra, C.: Review on computational trust and reputation models. Artificial Intelligence Review **24**(1) (September 2005) 33–60
2. Sierra, C., Debenham, J.: Information-based agency. In: Proceedings of Twentieth International Joint Conference on Artificial Intelligence IJCAI-07, Hyderabad, India (January 2007) 1513–1518
3. Arcos, J.L., Esteva, M., Noriega, P., Rodríguez, J.A., Sierra, C.: Environment engineering for multiagent systems. Journal on Engineering Applications of Artificial Intelligence **18** (2005)
4. Sierra, C., Debenham, J.: Information-based agency. In Veloso, M., ed.: Twentieth International Joint Conference on AI. (January 2007) In press
5. Sierra, C., Debenham, J.: The LOGIC Negotiation Model. In: Proceedings Sixth International Conference on Autonomous Agents and Multi Agent Systems AAMAS-2007, Honolulu, Hawai'i (May 2007) 1026–1033
6. Kalfoglou, Y., Schorlemmer, M.: IF-Map: An ontology-mapping method based on information-flow theory. In Spaccapietra, S., March, S., Aberer, K., eds.: Journal on Data Semantics I. Volume 2800 of Lecture Notes in Computer Science. Springer-Verlag: Heidelberg, Germany (2003) 98–127
7. Li, Y., Bandar, Z.A., McLean, D.: An approach for measuring semantic similarity between words using multiple information sources. IEEE Transactions on Knowledge and Data Engineering **15**(4) (July / August 2003) 871 – 882
8. MacKay, D.: Information Theory, Inference and Learning Algorithms. Cambridge University Press (2003)
9. Knoke, D., Yang, S.: Social Network Analysis. Sage Publications, Inc., California (2008)
10. Friedkin, N., Johnsen, E.: Social influence networks and opinion change. Advances in Group Processes **16** (1999) 1–29
11. Tutzauer, F.: Entropy as a measure of centrality in networks characterized by path-transfer flow. Social Networks **29**(2) (May 2007) 249–265
12. Cheeseman, P., Stutz, J.: On The Relationship between Bayesian and Maximum Entropy Inference. In: Bayesian Inference and Maximum Entropy Methods in Science and Engineering. American Institute of Physics, Melville, NY, USA (2004) 445 – 461
13. Paris, J.: Common sense and maximum entropy. Synthese **117**(1) (1999) 75 – 93
14. Peleg, S., Werman, M., Rom, H.: A unified approach to the change of resolution: Space and gray-level. IEEE Transactions on Pattern Analysis and Machine Intelligence **11**(7) (July 1989) 739—742
15. Yager, R.R.: On the determination of strength of belief for decision support under uncertainty — part i: generating strengths of belief. Fuzzy Sets and Systems **142**(1) (2004) 117–128