

Evaluating Trustworthiness from Past Performances: Interval-based Approaches¹

Jonathan Ben-Naim and Henri Prade

Institut de Recherche en Informatique de Toulouse (IRIT-CNRS)
Université de Toulouse, 118 route de Narbonne
F-31062 Toulouse Cedex 9, France
{bennaim, prade}@irit.fr

Abstract. In many multi-agent systems, the user has to decide whether he (or she) sufficiently trusts a certain agent to achieve a certain goal. To help users to make such decisions, an increasing number of trust systems have been developed. By trust system, we mean a system that gathers information about an agent and evaluates its trustworthiness from this information. The aim of the present paper is to develop new trust systems that overcome limitations of existing ones. This is a challenging problem that raises questions such as: how trustworthiness may be represented, and from which information it may be estimated? We assume that a set of grades describing the past performances of the agent is given. With this common basis, two approaches are proposed. In the first one, the aim is to construct an interval that summarizes the grades. Such an interval gives a good account of the trustworthiness of the agent. We establish axioms that should be satisfied by summarizing methods, devise a particular method based on pulling, and check that it satisfies the axioms, which provides theoretical justifications for it. In the second approach, which is more briefly presented, a level of trust as the certainty that a future grade will be good, and a level of distrust as the fear that a future grade may be bad, are computed on the basis of the past grades. This approach is based on possibility theory and provides, thanks to the two levels, another view of trustworthiness, as well as summarizing intervals.

Keywords: Trust, distrust, intervals, possibility theory.

1 Introduction

In many multi-agent systems, especially in the field of e-commerce, the user has a lot of decisions to make about the agents. For example, he (or she) has to decide whether he believes that an agent is more trustworthy than another, or whether an agent is sufficiently trustworthy in the absolute sense. This is the case for example in the famous auction system Ebay. Indeed, a buyer has to compare sellers and to decide whether he sufficiently trusts them to be honest

¹ The first author is supported by the French ANR-SETIN project ForTrust.

and competent. In order to help users to make such decisions, an increasing number of *trust systems* have been developed, see e.g. [SS05] for a review.

By trust system, we mean a system composed of two parts. First, an environment in which an agent called the trustee may be put as well as a certain kind of information about it. And second, a method which (in any possible situation) constructs from this information an object giving a good account of the trustworthiness of the trustee. This object is meant to provide trustors with some help to make decisions about trustees.

For example, Ebay is equipped with a system that takes a seller, collects the grades buyers gave to him, and attributes to him a certain number of “stars” on the basis of these grades. This number indicates how reliable the seller is, and provides buyers with a new criterion for judging and comparing sellers.

The aim of the present paper is to develop new trust systems which overcome limitations of existing ones. This is a challenging problem that raises questions such as: how trustworthiness may be represented, and from which information it may be estimated? We assume that a set of grades describing the past performances of the trustee is given. By convention, the value of a grade is a real number between 0 and 1. Each grade describes how perfectly the trustee achieved some past goal. Naturally, 0 means that it did not at all achieved it, whilst 1 means on the contrary that it achieved it perfectly.

With this common basis, two approaches are proposed. In the first one, the aim is to construct a *summarizing interval*, that is, an interval which necessarily describes the past behavior of the trustee with less accuracy than the grades, but gives a good account of the essential parts of this behavior. Such an interval constitutes a concise representation of the trustworthiness of the trustee. An advantage is that it is simpler than a set of grades. Therefore, it provides a new criterion for judging and comparing trustees, that is, a trustor may use it to evaluate its trust (or distrust) in the trustee, and thus to take decisions.

First, we establish intuitive and desirable properties, called axioms, which should be satisfied by *summarizing methods*. Next, we devise a particular method based on the idea that “strong” groups of grades have the strength to pull the bounds of the interval towards themselves. And finally, we check that our method satisfies the axioms, which provides theoretical justifications for it.

In certain cases, the user needs a *trust interval* rather than a summarizing one. By trust interval, we mean an interval such that it is rational to believe (on the basis of the past grades) that the future grades will essentially fall on it. We think that we get such an interval if we take a summarizing one and add an adequate margin of error. We provide a method for constructing such a margin. It is based on the fusion of the summarizing interval and the interval $[0, 1]$.

In the second approach, which is more briefly presented, the idea is to compute a *level of trust* as the certainty (on the basis of the past grades) that a future grade will be good, and a *level of distrust* as the fear that a future grade may be bad. These two levels provide another representation of the trustworthiness of the trustee. This approach is based on possibility theory, that is, we view the grades as a basis for building an histogram, which is then transformed into

a possibility distribution, which in turn provides a basis for computing levels of trust and distrust. Moreover, the possibility distribution so obtained may also be summarized under the form of a crisp interval by computing lower and upper expected values. This may be viewed as another way to build summarizing intervals. Besides, partial and complete orders are also proposed for comparing directly histograms of grade values, from a trustworthiness point of view.

The rest of the paper is structured as follows. In Section 2, we discuss different research trends on the problem of evaluating trustworthiness, and we highlight some advantages of our approaches. In Section 3, we present the first one, that is, in Section 3.1, we establish six axioms for summarizing methods, in Section 3.2, we develop a method based on pulling and prove that it satisfies the axioms, and in Section 3.3, we show how summarizing intervals may be used to construct trust intervals. Next, in Section 4, we turn to the second approach. We begin with histograms of grade values, introduce ordering relations between them, build possibility distributions from them, and finally obtain from these distributions levels of trust and distrust, as well as summarizing intervals.

2 Trust evaluation: different research trends

The problem of evaluating trustworthiness may be considered from many viewpoints. First of all, there are different kinds of representation formats. For example, trustworthiness may be represented by a number, a pair of numbers (usually one for trust and one for distrust), an interval, or even a fuzzy interval. In addition, the same representation format may have different understandings.

For example, in some approaches (e.g. [JK98] and [ZdSM05]), an agent is either trustworthy or not, and the authors manipulate a number which indicates the probability or the belief that it is trustworthy. In other approaches (e.g. [dCdS06]), an agent is trustworthy to a certain degree, and the authors work with a number that indicates to which degree. But, the probability of being trustworthy is certainly not the same thing as a degree of trustworthiness.

Other approaches motivate interval-based representation of trustworthiness by the poor quality of the information available, e.g. [Pra07]. This view is compatible with the understanding that trustworthiness is binary and the probability of being trustworthy is imprecise. Similarly, it is compatible with the understanding that trustworthiness is graded and the degree of trustworthiness is ill-known.

Another facet of the problem is that it may refer to quite different types of input data. For example, these data may take the form of opinions of agents about others, e.g. [AMT05], [WJI04], and [MDB06]. This is also the case in [TB04] and [TLU06], where trust evaluation is based on direct opinions, that is, opinions obtained from past interactions. Then, indirect opinions are computed by chaining and combining direct ones, either by means of inference rules [TLU06] or by path semirings [TB04]. Both approaches associate an uncertainty estimate to their trust values. Another option is to view trust assessment as a matter of argumentation, that is, the idea is to balance arguments in favor of deciding to trust an agent with arguments against this choice [Pra07].

In this paper, we evaluate trustworthiness from past performances, which has been rather neglected in the literature. Although purely statistical methods may be considered if enough data are available, we investigate other roads here since data are not necessarily numerous in practice. In e-commerce, there are systems (e.g. Ebay) where agents receive simple rates (e.g. good, bad, or neutral), and their trustworthiness, represented by a number, is evaluated in a rudimentary way from these rates. By our interval-based approaches, we try to exploit the rates in a deeper way. In addition, a number does not give many indications of how the rates are scattered, while an interval does.

3 Trustworthiness as a summarizing interval

3.1 Axioms

In this section, we provide desirable properties for summarizing methods.

Definition 1. A grade structure \mathcal{G} is an ordered pair $\langle G, v \rangle$, where G is a non-empty and finite set of grades and v is a function from a superset of G to the interval $[0, 1]$ (of real numbers). We call $v(g)$ the value of g .

A summarizing method I is a function which transforms any grade structure into a sub-interval of $[0, 1]$.

The reader may wonder why we allow the domain of v to be a superset of G , while G would be sufficient. The reason is simply that it helps to increase the readability of certain definitions and proofs. For all $x \in [0, 1]$, we denote by $wei_{\mathcal{G}}(x)$ the “weight” of x in \mathcal{G} , that is, $wei_{\mathcal{G}}(x) = |\{g \in G : v(g) = x\}|$. When the context is clear we may drop the subscript \mathcal{G} . The same goes for all notations.

We turn to a first obvious property. If two structures lead to the same values and to the same weights, then of course they should lead to the same interval.

Definition 2. Let $\mathcal{G} = \langle G, v \rangle$ and $\mathcal{G}' = \langle G', v' \rangle$ be two grade structures. We say that \mathcal{G} and \mathcal{G}' are equivalent (in symbols $\mathcal{G} \equiv \mathcal{G}'$) iff $v(G) = v'(G')$ and $\forall x \in v(G), wei_{\mathcal{G}}(x) = wei_{\mathcal{G}'}(x)$.

A summarizing method I respects equivalence iff for all grade structures \mathcal{G} and \mathcal{G}' , if $\mathcal{G} \equiv \mathcal{G}'$, then $I(\mathcal{G}) = I(\mathcal{G}')$.

Another obvious property is that the summarizing interval should not exceed the limits of the zone in which the grades are located.

Definition 3. A summarizing method I respects confinement iff for all grade structure \mathcal{G} , $I(\mathcal{G}) \subseteq [\min(v(G)), \max(v(G))]$.

Next, assume that the grades are regularly scattered over some distance. Then, the summarizing interval should cover exactly this distance.

Definition 4. A grade structure \mathcal{G} is regular iff:

- $\exists r, s \in \mathbb{R}, \exists n \in \mathbb{N}, v(G) = \{r, r + s, r + 2s, \dots, r + ns\}$
- $\forall x, y \in v(G), wei(x) = wei(y)$

A summarizing method I respects regularity iff for all regular grade structure \mathcal{G} , $I(\mathcal{G}) = [\min(v(G)), \max(v(G))]$.

The next property says the following: if the grades are symmetric with respect to some axis, then so should be the summarizing interval. Let $x, a \in \mathbb{R}$. We denote by $mir_a(x)$ is the mirror image of x with respect to a , that is, $mir_a(x) = 2a - x$.

Definition 5. A grade structure \mathcal{G} is symmetric with respect to $a \in \mathbb{R}$ iff for all $x \in v(\mathcal{G})$, $mir(x) \in v(\mathcal{G})$ and $wei(x) = wei(mir(x))$.

A summarizing method I respects symmetry iff for all grade structure \mathcal{G} and $\forall a \in \mathbb{R}$, if \mathcal{G} is symmetric with respect to a , then $mir(minI(\mathcal{G})) = maxI(\mathcal{G})$.

Note that we sometimes write $minX$, vX , etc. instead of $min(X)$, $v(X)$, etc. in order to increase readability. We denote by $mean_{\mathcal{G}}$ the mean of the grades of \mathcal{G} and by $cen_{\mathcal{G}}$ the center of the zone in which they are located, that is, $cen_{\mathcal{G}}$ is the middle of $[min(v(\mathcal{G})), max(v(\mathcal{G}))]$.

Assume that $mean$ is to the right of cen . Then, intuitively, \mathcal{G} is leaning to the right. The summarizing interval should reflect this asymmetry, that is, it should “forget” at least a bit the grades on the extreme left. The same goes when \mathcal{G} is leaning to the left.

Definition 6. A method I respects leaning iff for all grade structure \mathcal{G} ,

- if $cen < mean$, then $min(v(\mathcal{G})) < minI(\mathcal{G})$
- if $mean < cen$, then $maxI(\mathcal{G}) < max(v(\mathcal{G}))$

Finally, take a grade structure \mathcal{G} , add new grades to the right, and call \mathcal{G}' the structure so obtained. Then, the bounds of $I(\mathcal{G}')$ should be at least as to the right as those of $I(\mathcal{G})$. In addition, if some new grades are strictly more to the right than the old ones, then the right bound of $I(\mathcal{G}')$ should be strictly more to the right than that of $I(\mathcal{G})$. The same goes with left.

Definition 7. Let $\mathcal{G} = \langle G, v \rangle$ and $\mathcal{G}' = \langle G', v' \rangle$ be two grade structures. We denote by \preceq_r the relation such that

$$\mathcal{G} \preceq_r \mathcal{G}' \text{ iff } G' = G \cup H, \ v|_G = v'|_G, \text{ and } \forall g \in G, \forall h \in H, v'(g) \leq v'(h)$$

We denote by \prec_r the relation such that

$$\mathcal{G} \prec_r \mathcal{G}' \text{ iff } \mathcal{G} \preceq_r \mathcal{G}' \text{ and } \exists g' \in G', \forall g \in G, v(g) < v'(g')$$

The definitions of \preceq_l and \prec_l are obtained by replacing \leq by \geq , and $<$ by $>$.

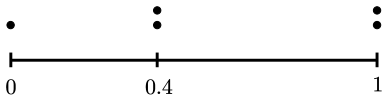
Intuitively, $\mathcal{G} \preceq_r \mathcal{G}'$ means that \mathcal{G}' can be obtained from \mathcal{G} by adding new grades to the right, and $\mathcal{G} \prec_r \mathcal{G}'$ means that some new grades are strictly more to the right than the old ones. Note that $v|_G$ denotes the restriction of v to G .

Definition 8. A method I is coherent iff for all grade structures \mathcal{G} and \mathcal{G}' ,

- if $\mathcal{G} \preceq_r \mathcal{G}'$, then $minI(\mathcal{G}) \leq minI(\mathcal{G}')$ and $maxI(\mathcal{G}) \leq maxI(\mathcal{G}')$
- if $\mathcal{G} \prec_r \mathcal{G}'$, then $maxI(\mathcal{G}) < maxI(\mathcal{G}')$
- if $\mathcal{G} \preceq_l \mathcal{G}'$, then $minI(\mathcal{G}') \leq minI(\mathcal{G})$ and $maxI(\mathcal{G}') \leq maxI(\mathcal{G})$
- if $\mathcal{G} \prec_l \mathcal{G}'$, then $minI(\mathcal{G}') < minI(\mathcal{G})$

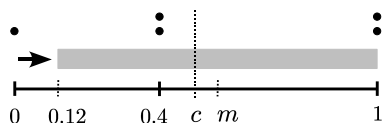
3.2 A summarizing method based on pulling

Assume a grade structure \mathcal{G} is given. It can be visualized as a set of points above the real interval $[0, 1]$, each point being located above its value. For example, the structure \mathcal{G}_0 consisting of 0, 0.4, 0.4, 1, and 1 can be visualized as follows:



We are now going to construct an interval $I_p(\mathcal{G})$ that summarizes \mathcal{G} . Initially, $I_p(\mathcal{G})$ is the smallest interval that contains the values of all points, which seems natural. Next, the idea is to look for “strong” groups of points, that is, groups able to pull the bounds of $I_p(\mathcal{G})$ towards themselves, despite the resistance of certain points. Finally, we move each bound of $I_p(\mathcal{G})$ to the farthest value x such that there exists a group able to pull it to x .

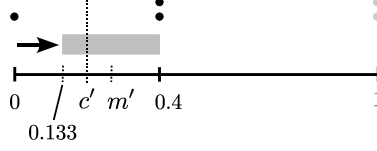
Here is a first simple way to find a group able to pull the left bound of $I_p(\mathcal{G})$ (the case of the right bound is similar). Suppose that the arithmetic mean m of the grades is to the right of the center c of $I_p(\mathcal{G})$. In our opinion, this means that the points above the right half of $I_p(\mathcal{G})$ constitute a group S able to pull the left bound towards the right. The more m is far from c , the more S is able to give it a hard pull. More precisely, S can pull the left bound “until c reaches m ”, that is, it can pull it to the value x such that if the left bound was equal to x , then c would be equal to m . For example, in \mathcal{G}_0 , the two 1’s are able to pull the left bound of $I_p(\mathcal{G}_0)$ to 0.12.



We turn to a more general way to find a group able to pull the left bound of $I_p(\mathcal{G})$. Take some limit l , ignore the points of \mathcal{G} to the right of l , and let \mathcal{G}' be the grade structure so obtained. Next, let $I_p(\mathcal{G}')$ be the smallest interval which contains the values of all grades of \mathcal{G}' , and, as before, let us look for a group able to pull the left bound of $I_p(\mathcal{G}')$. If we find such a group, then it is able to pull the left bound of $I_p(\mathcal{G})$ as well.

Indeed, suppose the mean m' of the grades of \mathcal{G}' is to the right of the center c' of $I_p(\mathcal{G}')$. Then, the points above the right half of $I_p(\mathcal{G}')$ constitute a group S' able to pull the left bound of $I_p(\mathcal{G}')$ until c' reaches m' . But, if a group of points can pull the left bound towards the right in a certain context, then of course it can do the same thing in any context obtained by adding new points to the right of this group. This reflects the intuition that he who can do more can do less. Consequently, since, in \mathcal{G}' , S' can pull the left bound of $I_p(\mathcal{G}')$ to a certain value x , it follows that, in \mathcal{G} , S' can pull the left bound of $I_p(\mathcal{G})$ to x as well.

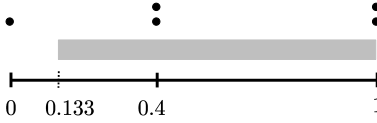
For example, let \mathcal{G}'_0 be the structure obtained from \mathcal{G}_0 by removing the two 1's. Then, in \mathcal{G}'_0 , the two 0.4's can pull the left bound of $I_p(\mathcal{G}'_0)$ to 0.133. Therefore, in \mathcal{G}_0 , the two 0.4's can pull the left bound of $I_p(\mathcal{G}_0)$ to the same value.



Consequently, in \mathcal{G}_0 , there are two groups able to pull the left bound of $I_p(\mathcal{G}_0)$, namely the two 1's (to 0.12) and the two 0.4's (to 0.133). It turns out that the second group is able to pull the left bound to a farther value than the first one. This may seem surprising since the first group is farther than the second one, but actually this is normal. Indeed, the two 1's are farther but they face a stronger resistance to pulling, that is, when pulling they are opposed to 0, 0.4, and 0.4, while the two 0.4's are only opposed to 0.

To summarize, we begin with the smallest interval $I_p(\mathcal{G})$ which contains the values of all grades of \mathcal{G} , then we identify the groups of grades able to pull the left bound of $I_p(\mathcal{G})$, and finally we move it to the farthest value x such that there exists a group able to pull it to x . The same goes for the right bound.

For example, here is in the final analysis the interval $I_p(\mathcal{G}_0)$ obtained from \mathcal{G}_0 by the pulling method (it is represented by the grey bar). Note that, according to our method, no group is able to pull the right bound of $I_p(\mathcal{G}_0)$.



Definition 9. Let $\mathcal{G} = \langle G, v \rangle$ be a grade structure. We denote by $I_p^s(\mathcal{G})$ the interval that superficially summarizes \mathcal{G} according to the pulling method, that is, the interval obtained by this method when no limit l is considered. More formally,

$$I_p^s(\mathcal{G}) = \begin{cases} [\min(vG), \max(vG)] & \text{if } cen = mean \\ [mir_{mean}(\max(vG)), \max(vG)] & \text{if } cen < mean \\ [\min(vG), mir_{mean}(\min(vG))] & \text{if } mean < cen \end{cases}$$

A set L is a *left part* of \mathcal{G} iff $\exists l \in \mathbb{R}, L = \{g \in G : v(g) \leq l\}$. Similarly, a set R is a *right part* of \mathcal{G} iff $\exists l \in \mathbb{R}, R = \{g \in G : l \leq v(g)\}$. We denote by $L(\mathcal{G})$ (resp. $R(\mathcal{G})$) the set of all non-empty left (resp. right) parts of \mathcal{G} .

Definition 10. Let $\mathcal{G} = \langle G, v \rangle$ be a grade structure. We denote by $I_p(\mathcal{G})$ the interval that summarizes \mathcal{G} according to the pulling method, that is,

$$I_p(\mathcal{G}) = [\max\{\min I_p^s\langle L, v \rangle : L \in L(\mathcal{G})\}, \min\{\max I_p^s\langle R, v \rangle : R \in R(\mathcal{G})\}]$$

Here are some additional examples to give the reader a better idea how I_p behaves. We scattered five grades regularly over some distance. Then, as desired, the interval obtained by the pulling method covers exactly the five grades.



Next, suppose that we bring closer together the three grades which are closest to the middle, while keeping a symmetric situation. Then, the newly formed concentration of grades makes us forget a bit the two extreme grades.



If we do the same thing, but without keeping a symmetric situation, then the concentration of grades makes us forget an extreme grade more than the other.



More formally, the summarizing method based on pulling satisfies all axioms introduced in Section 3.1.

Proposition 1. *The summarizing method I_p based on pulling satisfies equivalence, confinement, regularity, symmetry, leaning, and coherence.*

Proof. Equivalence, confinement, regularity, and leaning are easy.

Proof for symmetry. Assume $\mathcal{G} = \langle G, v \rangle$ is symmetric with respect to $a \in [0, 1]$. We show later that:

$$(0) \quad \forall L \in L(\mathcal{G}), \exists R \in R(\mathcal{G}), \text{mir}(\min I_p^s \langle L, v \rangle) = \max I_p^s \langle R, v \rangle$$

$$(1) \quad \forall R \in R(\mathcal{G}), \exists L \in L(\mathcal{G}), \text{mir}(\max I_p^s \langle R, v \rangle) = \min I_p^s \langle L, v \rangle$$

By definition, $\exists L \in L(\mathcal{G}), \min I_p(\mathcal{G}) = \min I_p^s \langle L, v \rangle$.

By (0), $\exists R \in R(\mathcal{G}), \text{mir}(\min I_p^s \langle L, v \rangle) = \max I_p^s \langle R, v \rangle$.

Therefore, $\max I_p(\mathcal{G}) \leq \max I_p^s \langle R, v \rangle = \text{mir}(\min I_p(\mathcal{G}))$.

We show $\text{mir}(\min I_p(\mathcal{G})) \leq \max I_p(\mathcal{G})$.

Suppose $\max I_p(\mathcal{G}) < \text{mir}(\min I_p(\mathcal{G}))$.

By definition, $\exists R \in R(\mathcal{G}), \max I_p(\mathcal{G}) = \max I_p^s \langle R, v \rangle$.

Thus, by (1), $\exists L \in L(\mathcal{G}), \text{mir}(\max I_p^s \langle R, v \rangle) = \min I_p^s \langle L, v \rangle$.

Thus, $\min I_p(\mathcal{G}) = \text{mir}(\text{mir}(\min I_p(\mathcal{G}))) < \text{mir}(\max I_p(\mathcal{G})) = \text{mir}(\max I_p^s \langle R, v \rangle) = \min I_p^s \langle L, v \rangle$, which is impossible.

Proof of (0) ((1) is similar). Let $L \in L(\mathcal{G})$. Then, $\exists l \in \mathbb{R}, L = \{g \in G : v(g) \leq l\}$.

Let $R = \{g \in G : \text{mir}(l) \leq v(g)\}$. We show $\text{mir}(v(L)) = v(R)$.

“ \subseteq ”. Let $y \in \text{mir}(v(L))$. Then, $\exists x \in v(L), y = \text{mir}(x)$. But, $x \leq l$.

Thus, $\text{mir}(l) \leq y$. In addition, by symmetry, $\exists g \in G, v(g) = y$.

But, $g \in R$. Therefore, $y \in v(R)$.

“ \supseteq ”. Let $x \in v(R)$. Then, $\text{mir}(l) \leq x$. Thus, $\text{mir}(x) \leq \text{mir}(\text{mir}(l)) = l$.

By symmetry, $\exists g \in G, v(g) = \text{mir}(x)$. But, $g \in L$. Thus, $\text{mir}(x) \in v(L)$.

Therefore, $x = \text{mir}(\text{mir}(x)) \in \text{mir}(v(L))$.

In addition, $L \neq \emptyset$. Thus, $R \neq \emptyset$. Therefore, $R \in R(\mathcal{G})$.

Let $c, m, \text{min}, \text{max}$ be shorthands for $\text{cen}_{\langle L, v \rangle}, \text{mean}_{\langle L, v \rangle}, \text{min}(v(L)), \text{max}(v(L))$.

We show $\text{mir}(c) = \text{cen}_{\langle R, v \rangle}$.

By definition, $\overrightarrow{c \text{min}} + \overrightarrow{c \text{max}} = \overrightarrow{0}$. But $\forall x, y \in \mathbb{R}, \overrightarrow{xy} = \overrightarrow{\text{mir}(y)\text{mir}(x)}$.

Therefore, $\overrightarrow{\text{mir}(\text{min})\text{mir}(c)} + \overrightarrow{\text{mir}(\text{max})\text{mir}(c)} = \overrightarrow{0}$.

But, $\text{mir}(v(L)) = v(R)$. So, $\text{mir}(\text{min}) = \text{max}(v(R))$ and $\text{mir}(\text{max}) = \text{min}(v(R))$.

Thus, $\overrightarrow{\text{max}(v(R))\text{mir}(c)} + \overrightarrow{\text{min}(v(R))\text{mir}(c)} = \overrightarrow{0}$. So, $\text{mir}(c) = \text{cen}_{\langle R, v \rangle}$.

We show $\text{mir}(m) = \text{mean}_{\langle R, v \rangle}$.

By definition, $\sum_{x \in v(L)} \overrightarrow{\text{wei}(x)\text{m}x} = \overrightarrow{0}$. Therefore,

$$\sum_{x \in v(L)} \overrightarrow{\text{wei}(x)\text{mir}(x)\text{mir}(m)} = \sum_{x \in v(L)} \overrightarrow{\text{wei}(\text{mir}(x))\overrightarrow{\text{mir}(x)\text{mir}(m)}} = \overrightarrow{0}.$$

Consequently, $\sum_{y \in v(R)} \overrightarrow{\text{wei}(y)y \text{mir}(m)} = \overrightarrow{0}$. Thus, $\text{mir}(m) = \text{mean}_{\langle R, v \rangle}$.

Finally, we show $\text{mir}_a(\text{min}I_p^s\langle L, v \rangle) = \text{max}I_p^s\langle R, v \rangle$.

Case 1: $c \leq m$. Then, $\text{min}I_p^s\langle L, v \rangle = \text{mir}_m(\text{max})$ and $\text{mir}_a(m) \leq \text{mir}_a(c)$.

Therefore, $\text{max}I_p^s\langle R, v \rangle = \text{mir}_{\text{mir}_a(m)}(\text{min}(vR)) = \text{mir}_{\text{mir}_a(m)}(\text{mir}_a(\text{max})) = \text{mir}_a(\text{mir}_m(\text{max}))$.

Case 2: $m < c$. Then, $\text{min}I_p^s\langle L, v \rangle = \text{min}$ and $\text{mir}_a(c) < \text{mir}_a(m)$.

Thus, $\text{max}I_p^s\langle R, v \rangle = \text{max}(v(R)) = \text{mir}_a(\text{min})$.

Proof for coherence. Let $\mathcal{G} = \langle G, v \rangle$ and $\mathcal{G}' = \langle G', v' \rangle$ be two grade structure.

Suppose $\mathcal{G} \preceq_r \mathcal{G}'$. We show $\text{min}I_p(\mathcal{G}) \leq \text{min}I_p(\mathcal{G}')$.

There exists $L \in L(\mathcal{G})$ such that $\text{min}I_p(\mathcal{G}) = \text{min}I_p^s\langle L, v \rangle$.

Let $L' = \{g \in G' : v'(g) \leq \text{max}(vL)\}$. Then, $L' \in L(\mathcal{G}')$.

Let m, m', c, c' be shorthands for $\text{mean}_{\langle L, v \rangle}, \text{mean}_{\langle L', v' \rangle}, \text{cen}_{\langle L, v \rangle}, \text{cen}_{\langle L', v' \rangle}$.

Then, $\text{min}(vL) = \text{min}(v'L')$, $\text{max}(vL) = \text{max}(v'L')$, $c = c'$ and $m \leq m'$.

Case 1: $c \leq m$ and $c' \leq m'$.

Then, $\text{min}I_p^s\langle L, v \rangle = \text{mir}_m(\text{max}(vL)) \leq \text{mir}_{m'}(\text{max}(v'L')) = \text{min}I_p^s\langle L', v' \rangle$.

Case 2: $c \leq m$ and $m' < c'$. Then, $m' < m$, which is impossible.

Case 3: $m < c$ and $c' \leq m'$.

Then, $\text{min}I_p^s\langle L, v \rangle = \text{min}(vL) = \text{min}(v'L') \leq \text{mir}_{m'}(\text{max}(v'L')) = \text{min}I_p^s\langle L', v' \rangle$.

Case 4: $m < c$ and $m' < c'$.

Then, $\text{min}I_p^s\langle L, v \rangle = \text{min}(vL) = \text{min}(v'L') = \text{min}I_p^s\langle L', v' \rangle$.

We show $\text{max}I_p(\mathcal{G}) \leq \text{max}I_p(\mathcal{G}')$.

There exists $R' \in R(\mathcal{G}')$ such that $\text{max}I_p(\mathcal{G}') = \text{max}I_p^s\langle R', v' \rangle$.

Let $R = \{g \in G : \text{min}(v'R') \leq v(g)\}$.

Case 1: $R = \emptyset$. Then, $\text{max}I_p(\mathcal{G}) \leq \text{max}(vG) < \text{min}(v'R') \leq \text{max}I_p(\mathcal{G}')$.

Case 2: $R \neq \emptyset$. Then, $R \in R(\mathcal{G})$. Thus, $\text{max}I_p(\mathcal{G}) \leq \text{max}I_p^s\langle R, v \rangle$.

Let m, m', c, c' be shorthands for $\text{mean}_{\langle R, v \rangle}, \text{mean}_{\langle R', v' \rangle}, \text{cen}_{\langle R, v \rangle}, \text{cen}_{\langle R', v' \rangle}$.

Then, $\text{min}(vR) = \text{min}(v'R')$, $\text{max}(vR) \leq \text{max}(v'R')$, and $m \leq m'$.

Case 2.1: $c \leq m$ and $c' \leq m'$.

Then, $\text{max}I_p^s\langle R, v \rangle = \text{max}(vR) \leq \text{max}(v'R') = \text{max}I_p^s\langle R', v' \rangle$.

Case 2.2: $c \leq m$ and $m' < c'$. Then, $\text{max}I_p^s\langle R, v \rangle = \text{max}(vR) \leq \text{mir}_m(\text{min}(vR))$

$\leq \text{mir}_{m'}(\text{min}(v'R')) = \text{max}I_p^s\langle R', v' \rangle$.

Case 2.3: $m < c$ and $c' \leq m'$.

Then, $\max I_p^s \langle R, v \rangle = \min(\min(vR)) < \max(vR) \leq \max(v'R') = \max I_p^s \langle R', v' \rangle$.

Case 2.4: $m < c$ and $m' < c'$.

Then, $\max I_p^s \langle R, v \rangle = \min(\min(vR)) \leq \min(\min(v'R')) = \max I_p^s \langle R', v' \rangle$.

Suppose $\mathcal{G} \prec_r \mathcal{G}'$. We show $\max I_p(\mathcal{G}) < \max I_p(\mathcal{G}')$.

The proof is similar to that of $\max I_p(\mathcal{G}) \leq \max I_p(\mathcal{G}')$.

The difference is that this time we have $\max(vR) < \max(v'R')$ and $m < m'$.

This difference allows us to derive in all cases $\max I_p^s \langle R, v \rangle < \max I_p^s \langle R', v' \rangle$.

The proofs for \preceq_l and \prec_l and similar to those for \preceq_r and \prec_r . \square

3.3 From summarizing intervals to trust intervals

In certain contexts, given a grade structure $\mathcal{G} = \langle G, v \rangle$, the user needs a trust interval rather than a summarizing one. Recall that a trust interval is an interval such that it is rational to believe (on the basis of the past grades) that the future grades will essentially fall on it. We think that we get such an interval $T(\mathcal{G})$ if we take a summarizing interval $I(\mathcal{G})$ and add an adequate margin of error.

The question is of course: what is an adequate margin of error? The more the number of past grades is big, the more the margin should be small. A solution is for example to define the left bound of $T(\mathcal{G})$ as the weighted mean of 0 and the left bound of $I(\mathcal{G})$. The weight of the former is 1, while that of the latter is the number of past grades, that is, $|G|$. The same goes for the right bound.

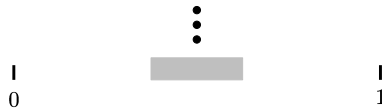
Definition 11. Let $\mathcal{G} = \langle G, v \rangle$ be a grade structure. We denote by $T_p(\mathcal{G})$ the trust interval obtained from \mathcal{G} according to the pulling method, that is,

$$T_p(\mathcal{G}) = \left[\frac{|G| \min I_p(\mathcal{G})}{1 + |G|}, \frac{1 + |G| \max I_p(\mathcal{G})}{1 + |G|} \right]$$

T_p and I_p behave almost in the same way. The essential difference is that two structures may lead to the same summarizing interval, but to different trust intervals. This is possible because there exist structures which contain different numbers of grades, and yet lead to the same summarizing interval. Here is an example. Suppose that there is only one grade of value 0.5. Then, T_p is centered and strictly smaller than $[0, 1]$, which reflects the idea that we are expecting future grades close to 0.5, but we remain cautious.



Next, add two other 0.5's. Then, T_p is still centered and even smaller than before (as desired), while in both cases I_p is the same interval, namely $[0.5, 0.5]$.



Provided that our view of a trust interval is convincing (“trust = summary + margin of error”), T_p provides trustors with an indication of what they can expect from trustees, which may be of a certain help.

4 Trustworthiness as levels of trust and distrust

Recall that the information available about the past behavior of an agent is supposed to take the form of a collection of grades whose values belong to the interval $[0, 1]$. The interval $[0, 1]$ is viewed in this section as a bipolar univariate scale. By convention, 0 is supposed to stand for the worst grade, and 1 for the best one. The value 0.5 will be considered as a neutral value, which means that values in $(0.5, 1]$ will be understood as increasingly good, while values in $[0, 0.5)$ are bad values, which are all the worse as they are closer to 0. Although, it may look natural, the choice of the neutral value in the middle of the interval is not compulsory, and the approach described in the following could be easily adapted to a neutral value differently located in $(0, 1)$.

We also assume that $[0, 0.5)$ (resp. $(0.5, 1]$) is partitioned into a finite number of subintervals of equal length, increasingly ordered $\lambda_{-n}, \dots, \lambda_{-j}, \dots, \lambda_{-1}$ (resp. $\mu_{+1}, \dots, \mu_{+i}, \dots, \mu_{+n}$). All grade values belonging to the same interval λ_{-j} (or μ_{+i}) are considered as indistinguishable and equivalent. Thus, a grade structure $\mathcal{G} = \langle G, v \rangle$ will be characterized by the sequence $\langle w_{-n}, \dots, w_{-1}, w_{+1}, \dots, w_{+n} \rangle$, where $w_{-j} = |\{g \in G : v(g) \in \lambda_{-j}\}|$ and similarly $w_{+i} = |\{g \in G : v(g) \in \mu_{+i}\}|$.

Let us first provide some examples of partial orders for comparing two grade structures. Given two structures $\mathcal{G} = \langle G, v \rangle$ and $\mathcal{G}' = \langle G', v' \rangle$, with respective sequences $\langle w_k \rangle$ and $\langle w'_k \rangle$, a partial order can be defined by comparing the global amounts of grades with good and with bad values. Namely,

$$\mathcal{G} \gg_{card} \mathcal{G}' \text{ iff } \sum_{i=1}^n w_{+i} \geq \sum_{i=1}^n w'_{+i} \text{ and } \sum_{j=1}^n w_{-j} \leq \sum_{j=1}^n w'_{-j}$$

Clearly, $\mathcal{G} \gg_{card} \mathcal{G}'$ expresses that the grade structure \mathcal{G} is at least as good (from a trustworthiness viewpoint) as \mathcal{G}' since \mathcal{G} has more grades with good values and less grades with bad values in the wide sense than \mathcal{G}' . An even more refined, actually complete, order can then be obtained by taking the difference of the global amounts of grades with good values and bad values.

$$\mathcal{G} \gg_{d-card} \mathcal{G}' \text{ iff } \sum_{i=1}^n w_{+i} - \sum_{j=1}^n w_{-j} \geq \sum_{i=1}^n w'_{+i} - \sum_{j=1}^n w'_{-j}$$

However, it makes no difference between a structure with a few good grade values and no bad ones, and a structure with some bad ones and a few more good ones. Moreover, it is clear that all the grades with good (resp. bad) values have not the same importance. A value is all the better (resp. worse) as it is closer to 1 (resp. 0). So, it would be advisable to use fuzzy cardinality [DP80] in the above definition. Let us first define the fuzzy set *Good* of good value classes and the fuzzy set *Bad* of bad value classes. We take for the membership degrees increasing and decreasing functions respectively:

$$\forall j, Good(\lambda_{-j}) = 0, \quad Good(\mu_{+1}) = \frac{1}{n}, \dots, Good(\mu_{+i}) = \frac{i}{n}, \dots, Good(\mu_{+n}) = 1$$

$$Bad(\lambda_{-n}) = 1, \dots, Bad(\lambda_{-j}) = \frac{j}{n}, \dots, Bad(\lambda_{-1}) = \frac{1}{n}, \quad \forall i, Bad(\mu_{+i}) = 0$$

For convenience, let us use the simplified notations $Good(\mu_{+i}) = go_{+i}$ and $Bad(\lambda_{-j}) = ba_{-j}$. Then, a fuzzy cardinality-based partial ordering is defined:

$$\mathcal{G} \gg_{f-card} \mathcal{G}' \text{ iff } \sum_{i=1}^n go_{+i} w_{+i} \geq \sum_{i=1}^n go_{+i} w'_{+i} \text{ and } \sum_{j=1}^n ba_{-j} w_{-j} \leq \sum_{j=1}^n ba_{-j} w'_{-j}$$

Such a comparison between two structures \mathcal{G} and \mathcal{G}' is fair only if they are based on the same number of grades, that is, only if $|G| = |G'|$. Otherwise, if we compare two structures with unequal number of grades, it may be better to use relative cardinality, that is, $\mathcal{G} \gg_{rf-card} \mathcal{G}'$ iff

$$\frac{1}{|G|} \sum_{i=1}^n go_{+i} w_{+i} \geq \frac{1}{|G'|} \sum_{i=1}^n go_{+i} w'_{+i} \text{ and } \frac{1}{|G|} \sum_{j=1}^n ba_{-j} w_{-j} \leq \frac{1}{|G'|} \sum_{j=1}^n ba_{-j} w'_{-j}$$

Let us illustrate on an example the different discriminating powers of the orderings \gg_{f-card} and $\gg_{rf-card}$. Let \mathcal{G}^1 , \mathcal{G}^2 , and \mathcal{G}^3 be three structures defined by:

- $w_{-n}^1 = 3$ and $\forall i, \forall j \neq n, w_{-j}^1 = w_{+i}^1 = 0$
- $w_{-n}^2 = 2$ and $\forall i, \forall j \neq n, w_{-j}^2 = w_{+i}^2 = 0$
- $w_{-n}^3 = 2, w_{-(n-1)}^3 = 1$ and $\forall i, \forall j < n-1, w_{-j}^3 = w_{+i}^3 = 0$

Then it can be checked that $\mathcal{G}^2 \gg_{f-card} \mathcal{G}^1$, while $\mathcal{G}^2 \approx_{rf-card} \mathcal{G}^1$ (where $a \approx b$ iff $a \gg b$ and $b \gg a$). However $\mathcal{G}^2 \gg_{f-card} \mathcal{G}^3$ and $\mathcal{G}^3 \gg_{rf-card} \mathcal{G}^2$. As expected, $\gg_{rf-card}$ is only sensitive to the percentages of rather good and rather bad cases, but not directly to their actual cardinalities. However, it acknowledges the fact that in average \mathcal{G}^3 is not as bad as \mathcal{G}^2 , although \mathcal{G}^2 reports less bad cases.

We now discuss the computation of a level of trust as the certainty (on the basis of the past grades) that a future grade will be good, and a level of distrust as the fear that a future grade may be bad. The idea is that the level of trust should be high if *always* grades with good values are reported, while the level of distrust should be high as soon as *some* grades with bad values are reported.

For computing such levels of trust and distrust, we use a two-step approach: i) transform the grade structure \mathcal{G} into a possibility distribution, then viewed as restricting the possible value of the next outcome; ii) on this basis, compute the level of trust as the necessity measure that a grade with a good value will be obtained, and the level of distrust as the possibility that a bad grade value will be obtained, where “good” and “bad” refer to the fuzzy sets *Good* and *Bad*.

For performing the first step, we first normalize the weighting structure, and then apply a probability-possibility transformation, preserving as much information as possible. For $j = 1, \dots, n$, let $p_{-j} = \frac{w_{-j}}{|G|}$. For $i = 1, \dots, n$, let $p_{+i} = \frac{w_{+i}}{|G|}$. Then, in order to apply the probability-possibility transformation [DPS93], we have to re-order the set $\{p_{-n}, \dots, p_{-1}, p_{+1}, \dots, p_{+n}\}$ decreasingly, as $p_{\sigma(1)} \geq \dots \geq p_{\sigma(k)} \geq \dots \geq p_{\sigma(2n)}$. Next, let $\pi_{\sigma(k)} = \sum_{j=k}^{2n} p_{\sigma(j)}$. Note that

$\pi_{\sigma(1)} \geq \dots \geq \pi_{\sigma(k)} \geq \dots \geq \pi_{\sigma(2n)}$, which expresses that the transformation is faithful with respect to the shape of the distribution.

The second step amounts to the computation of the necessity and the possibility of fuzzy events, that is,

$$\text{trust}(\mathcal{G}) = \min_x \max(\text{Good}(x), 1 - \pi(x)) \quad \text{distrust}(\mathcal{G}) = \max_x \min(\text{Bad}(x), \pi(x))$$

The level of distrust is high as soon as there exists a really bad grade value that is highly plausible. The level of trust is high as soon as any bad grade value (including the less bad ones) is impossible or almost impossible. Such definitions acknowledge the fact that one should be afraid by bad performances in trust evaluation, which play a more important role than the good ones. It can be checked that the sum of these two levels is always less or equal to 1 (as in most models of trust and distrust, e.g. [dCdS06], however these levels are computed here from past performances, good or bad, in agreement with possibility theory). Thus, from them an interval pertaining to trust (resp. to distrust) can be built as $[\text{trust}(\mathcal{G}), 1 - \text{distrust}(\mathcal{G})]$ (resp. $[\text{distrust}(\mathcal{G}), 1 - \text{trust}(\mathcal{G})]$). These levels (and intervals) not only involve the grade structure information \mathcal{G} , but also a graded view of goodness and badness.

Besides, the obtained possibility distribution π , or more precisely its convex hull, can be seen as a kind of fuzzy version of summarizing interval in the sense of Section 3.2. This interval may be transformed into a crisp one representing its mean value [DP87]. The bounds of the interval are computed as lower and upper expected values $E_*(\pi)$ and $E^*(\pi)$ (using Choquet integrals)

$$E_*(\pi) = \sum_{k=1}^n x_k(\pi_*(x_k) - \pi_*(x_{k-1})) \quad E^*(\pi) = \sum_{k=1}^n x_k(\pi^*(x_k) - \pi^*(x_{k+1}))$$

where $\pi_*(x) = \max_{t \leq x} \pi(t)$ and $\pi^*(x) = \max_{t \geq x} \pi(t)$ and the x_k stand for the (increasingly ordered) central values of the $2n$ subintervals partitioning $[0, 1]$. It can be shown that the interval $[E_*(\pi), E^*(\pi)]$ satisfies all the axioms of the first approach, presented in Section 3.1, with the exception of the leaning axiom.

Using the possibility distribution π , one may also compute the largest value \underline{x} such that $N(\underline{x} \leq) = 1$ or $\theta \leq N(\underline{x} \leq)$ and the smallest \bar{x} such that $\Pi(\bar{x} \leq) = 0$ or $\Pi(\bar{x} \leq) \leq \rho$, where θ and ρ are thresholds, Π and N are possibility and necessity measures with $\Pi(A) = \max_x \min(A(x), \pi(x))$ and $N(A) = 1 - \Pi(\neg A)$. This is again a summarizing interval $[\underline{x}, \bar{x}]$ (now based on Sugeno integrals).

5 Conclusion

The first contribution of the present paper is that we propose interval-based representations of trustworthiness (i.e. summarizing and trust intervals), which is rather new in the field of trust and reputation. The second contribution is that we provide two approaches for evaluating trustworthiness from a set of past performances, which has also been rather neglected in the literature. Interestingly enough, the first approach is based on a purely horizontal view (based on

the idea of pulling values), while the second approach exploits a vertical view (based on histograms, then transformed into possibility distributions).

Besides, general partial orders can be defined for comparing two grade structures, for example, the bipolar ones introduced at the beginning of Section 4, or unipolar ones such as stochastic dominance. This could be the basis of postulates expressing some agreement between partial orderings of grade structures and comparisons of their associated summarizing intervals.

Lines for further research also include a deeper comparison of the two approaches, the validation of the models from a cognitive psychology point of view, and refinements in order to take into account the freshness of the information.

References

- [AMT05] P. Avesani, P. Massa, and R. Tiella. A trust-enhanced recommender system application: Moleskiing. In *Proc. of the 2005 ACM symposium on Applied computing*, pages 1589–1593, 2005.
- [dCdS06] M. de Cock and P. Pinheiro da Silva. A many valued representation and propagation of trust and distrust. *Lecture Notes in Artificial Intelligence*, 3849:114–120, 2006.
- [DP80] D. Dubois and H. Prade. *Fuzzy sets and systems: Theory and applications*. Academic Press, 1980.
- [DP87] D. Dubois and H. Prade. The mean value of a fuzzy number. *Fuzzy Sets and Systems*, 24:279–300, 1987.
- [DPS93] D. Dubois, H. Prade, and S. Sandri. On possibility/probability transformations. In M. Roubens R. Lowen, editor, *Fuzzy logic: State of the art*, pages 103–112. Kluwer Academic Publ., 1993.
- [JK98] A. Jøsang and S. J. Knapskog. A metric for trusted systems. In *Proc. of the 21st NIST-NCSC National Information Systems Security Conference*, pages 16–29, 1998.
- [MDB06] D. Melaye, Y. Demazeau, and T. Bouron. Which adequate trust model for trust networks? In *Proc. of the 3rd IFIP Conference on Artificial Intelligence Applications and Innovations (AIAI)*, pages 236–244, 2006.
- [Pra07] H. Prade. A qualitative bipolar argumentative view of trust. *Lecture Notes in Computer Science*, 4772:268–276, 2007.
- [SS05] J. Sabater and S. Sierra. Review on computational trust and reputation models. *Artificial Intelligence*, 24:33–60, 2005.
- [TB04] G. Theodorakopoulos and J. S. Baras. Trust evaluation in ad-hoc networks. In *Proc. of the 3rd ACM workshop on Wireless security*, pages 1–10, 2004.
- [TLU06] S. Toivonen, G. Lenzini, and I. Uusitalo. Context-aware trustworthiness evaluation with indirect knowledge. In *Proc. of the 2nd International Semantic Web Policy Workshop (SWPW)*, 2006.
- [WJI04] A. Whitby, A. Jøsang, and J. Indulska. Filtering out unfair ratings in bayesian reputation systems. In *Proc. of the 7th Int Workshop on Trust in Agent Societies, at AAMAS'04*, 2004.
- [ZdSM05] I. Zaihrayeu, P. Pinheiro da Silva, and D. L. McGuinness. IWTrust: Improving user trust in answers from the web. *Lecture Notes in Computer Science*, 3477:384–392, 2005.