Introducing the *ImRep* model

Jordi Sabater, Mario Paolucci, Rosaria Conte
LABSS (Laboratory of Agent Based Social Simulation)
Institute of Cognitive Science and Technology
National Research Council
Viale Marx 15, 00137 Roma, Italy

Abstract

The area of computational trust and reputation models has evolved very quickly the last few years. Now we have a clearer understanding of the mechanisms that are behind trust and reputation. The models start to consider different aspects that influence trust and reputation improving their accuracy under complex environments. However the introduction of these new aspects arises a new problem. How the different pieces of knowledge have to be combined to get a final trust or reputation value? In this article we introduce ImRep*, a trust and reputation model that is a first attempt to undertake this problem from a cognitive perspective.*

1 Introduction

One of the main problems you can find in current trust and reputation models is what we call the "'aggregation problem"'. It is clear there are several elements, usually associated to different sources of information, that contribute to build trust and reputation. Different authors propose different divisions [7, 11, 1, 3] but sooner or later all of them arrive to the same point: it is necessary to combine the basic elements to get a final conclusion that can contribute to the decision making process.

The computational approach to this problem usually has been the use of a weighted mean. Each element is weighted by hand or following fixed heuristics that give more relevance to direct experiences, that use some kind of reliability measure as a value for the weight, and so on.

This approach can be enough for those situations where the environment is very well known and relatively static and therefore it is possible to tune the parameters of the model before entering into the scenario. However if we want a general method to combine the elements that compound trust and reputation, that can be used in different environments and that can adapt to changes, this approach is clearly insufficient.

Based on our previous experience trying to solve this problem [10], we think the only way to undertake it is using a cognitive approach.

In this article we present our ideas about the interplay of Image and Reputation (section 2) and then we introduce *ImRep*, a computational model that has been designed having in mind the problem of aggregating the elements used to build trust and reputation (section 3). Although we are still at an initial stage, we will present a general view of the model and we will show how it can be used to model the interplay of Image and Reputation.

2 Combining Image and Reputation

To analyze the interplay of Image and Reputation the first thing we have to do is to define what we mean by Image and Reputation.

Image is an evaluative belief; it tells whether the target is "good" or bad with respect to a given behaviour. An agent has an evaluation when he or she believes that a given entity is good for, or can achieve, a given goal. An agent has a social evaluation when his or her belief concerns another agent as a means for achieving this goal [2].

We represent the Image of an agent t (that we call the *target*) from the point of view of and agent s (that we call the *source*) related to aspect φ as $Image(s, t, \varphi, \alpha)$, where α is the value of the Image.

Our definition of reputation moves it one level above the image level. Reputation is a belief about others' minds, or more specifically about others' evaluations of the target; it is a meta-belief. This definition has one important consequence. To accept a meta-belief does not imply the acceptance of the nested belief. To assume that a target t is assigned a given reputation implies assuming that t is believed to be "good" or "bad," but it does not imply sharing either evaluation [2].

We represent the Reputation of an agent t from the point of view of an agent s related to aspect φ as $Rep(s, t, \varphi, \alpha)$, where α is the value of that Reputation.

Image and Reputation are two of the most (if not the most) relevant factors used to build trust. A lot of trust models (with slight variations, specially in the definition of Image) agree on that. This is why we have chosen the interplay of Image and Reputation as our first objective.

The situation we want to study is when the individual has an Image and a Reputation associated to a given target. Specifically we want to study the case when and individual s has $Image(s, t, \varphi, \alpha_I)$ and $Rep(s, t, \varphi, \alpha_R)$.

We will assume that $\alpha_I, \alpha_R \in \{Good, Bad\}$. Given that table 1 shows the different cases we have to consider.

	α_I	α_R
case-1	Good	Good
case-2	Good	Bad
case-3	Bad	Good
case-4	Bad	Bad

Table 1. Interplay of Image and Reputation

Cases 1 and 4 imply a stable mental state. The Image and the Reputation are coincident so from the point of view of the interplay, Reputation reinforces Image and vice versa. Given that Image and Reputation are based on different sources of information, we can say there is a *certainty* on the target about the aspect φ.

Cases 2 and 3, on the other side, imply a contradiction between two pieces of information that refer to the same aspect. This is what is called a *cognitive dissonance*. A *cognitive dissonance* generates an instability in the mind of the individual. Depending on how strong and relevant is the *cognitive dissonance*, the individual is pushed to solve it by taking special actions. Although these actions are context dependent, they are always oriented to confirm the grounds of the elements that are causing the *cognitive dissonance*. The question is: why these actions are not taken before arriving to a *cognitive dissonance*, that is, during the process of building each element? The reason is that these special actions usually imply a great cost in terms of resources (time, money, etc.). Given that, individuals adopt a lazy approach, only when it is really needed the individual will carry on the special actions to solve a *cognitive dissonance*.

Returning to the interplay of Image and Reputation, a *cognitive dissonance* appears when there is a contradiction in the agent's mind between a target image and the reputation of that target. Solving a *cognitive dissonance* in this case implies to check in deep the reliability of the Image and the reliability of the Reputation trying to find which one is wrong.

As we have said, the elements used to build an Image are direct experiences. There are three questions an individual has to answer:

- Is the number of direct experiences enough to infer that Image?
- Are the direct experiences outdated?
- Is there noise in the perception of the direct experiences?

In the case of Reputation, because it is build from a set of third party Images the questions to be answered are:

- Is the number of witnesses enough to infer a Reputation?
- Is there a problem of "'correlated evidence"'? (We talk about "'correlated evidence"' when the opinions of different witnesses are based on the same event(s) or when there is a considerable amount of shared information that tends to unify the witnesses' way of "thinking").
- Are the witnesses lying (what we call the "'complot theory"')?

Finally, taking into account both, Image and Reputation, there is the possibility that the criteria used to judge the direct experiences be different to that used by the rest of the community.

After trying to answer these questions there are several possibilities:

- The Image or the Reputation value was wrong. If one of the two elements is no longer true, the *cognitive dissonance* disappear.
- There is not enough available information to answer the questions. This usually implies the "'quarantine"' of the problem until there is more information available.
- The individual arrives to the conclusion that both, the Image and the Reputation values are apparently correct. Everything seems to give support to both elements but they are still contradictory.

There are psychological studies that demonstrate there is a bias toward bad Image and Reputation [12, 8], that is, individuals tend to give more relevance to bad Image and Reputation. This asymmetry will affect the behavior of the individual in the last two possibilities where there is no way to give predominance to one of the elements.

3 The *ImRep* model

In this section we will introduce *ImRep*, a trust and reputation model that has been specially designed to deal with the problem of aggregating the elements that are used to

build trust and reputation using a cognitive approach. We will use the interplay of Image and Reputation as the case study to present the model.

The model is divided in three parts: a memory with all the predicates that represent the knowledge that is associated to image and reputation, a set of processes we call *detectors* that operate on this memory; and finally a decision maker mechanism.

3.1 The agent architecture

Before starting the description of the elements that compound *ImRep* it is important to situate the model in the context of a generic agent architecture. In this section we describe those elements of an agent architecture that are relevant and necessary from the point of view of an *ImRep* model. As shown in figure 1, there are three elements that have to be considered: the agent memory, a planner and the communication module.

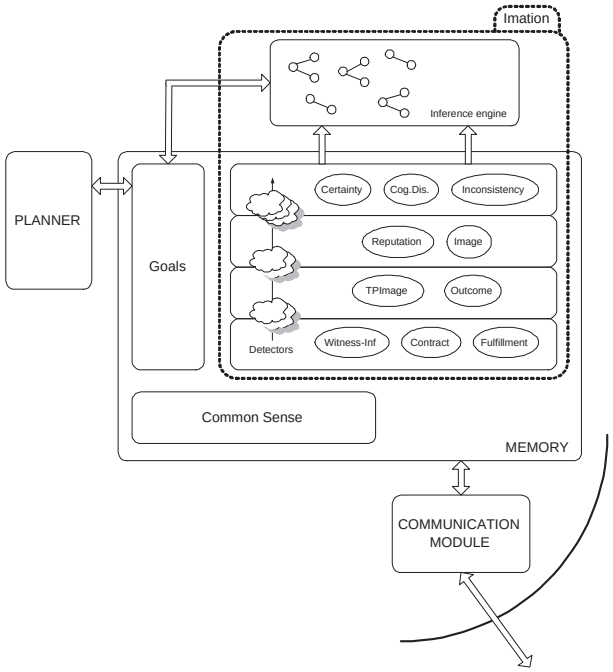

Figure 1. The *ImRep* model and its environment

The main memory is the most important part of the agent architecture from the point of view of the *ImRep* model. *ImRep* uses that memory to store the knowledge associated to the concepts of Image and Reputation.

From now on, we will assume that the main memory of the agent is a data base of first order predicates. Is in this memory where he agent stores its goals. The agent goals can be either those initially assigned by the user or those generated by a planner as intermediate steps to achieve the main goals. We assume also that each goal has an associated priority that determines its relevance in the context of the final objectives. These priorities are not static but can change along time depending on several factors like the environment or the agent mental status.

How the planner works as well as how the agent assigns priorities to the goals is something that goes beyond the scope of this article.

The goals' section in the main memory is reduced to a list of predicates that represent the pending goals, each one with an assigned priority. The priority goes from 0 (minimum priority) to 1 (maximum priority). A goal is represented as:

$$Goal(X, \phi, p)$$

where X is the agent that has to achieve the goal ϕ, and p is the priority that the agent assigns to this goal.

The agent goals and how they are prioritized influence a lot the way the *ImRep* model works. Surprisingly, in spite of its relevance, this link with the goals of the agent is usually ignored by current trust and reputation models.

The last element of the agent architecture we need to consider is the communication module. The communication module receives messages from the outside and adds the content into the memory in a suitable format. From the point of view of the *ImRep* model the relevant information is the information associated to contracts and its results as well as third party opinions about other agents.

3.2 The memory

As we said before, *ImRep* uses the main memory of the agent to store the knowledge associated to Image and Reputation. This knowledge is represented by first order predicates.

The predicates are conceptually organized in different levels as shown in figure 1. A predicate in a given level is always inferred from predicates that belong to the same or a lower level.

The lowest memory level is used to store the information that still has not been evaluated by the agent. There are three types of information at this level:

- Contracts: In this context, a contract is not necessary a formal contract. It can be just an agreement between two agents. It is represented as $Contract(s, t, I, X^c)$ where s and t are the agents involved in the contract, I a set of indexes that identify the issues of the contract and X^c a vector with the agreed values of the contract.
- Fulfillments: Is the result of a contract. For instance, in an e-commerce environment can be the real characteristics of a product that the agent bought several days

ago. It is represented as $Fulfillment(s, t, I, X^f)$ where s and t are the agents involved in the contract, I a set of indexes that identify the issues of the contract and X^f a vector with the actual values after the fulfillment of the contract.

- Witness information: information about the agent community coming from third party agents (witnesses). We represent witness information as $WInf(s, t, Content)$, that is, agent s informs agent t that $Content$ is true. In our case, $Content$ can be an image or a reputation associated to a third party agent.

The next level is populated by *outcomes* and *third party images*.

We define the *outcome* of a dialogue between two agents as:

- An initial contract to take a particular course of action and the actual result of the actions taken.
- An initial contract to fix the terms and conditions of a transaction and the actual values of the terms of the transaction.

An outcome is not just the tuple contract-fulfillment, is also the evaluation of this tuple considering how was fulfilled the contract. An outcome is represented with a predicate of the form $Outcome(s, t, I, X^c, X^f, \phi)$ where s and t are the agents involved in the contract, I a set of indexes that identify the issues of the contract, X^c and X^f are the vectors with the agreed values of the contract and the actual values after its fulfillment respectively and $\phi \in [0, 1]$ is the value of the evaluation made by the agent being $\phi = 0$ the worst and $\phi = 1$ the best.

A third party image ($TPImage$) is the evaluation that and agent makes in terms of the reliability of an image that a witness has on a target agent (received as the content of a witness information). This evaluation is based on a set of factors like the accuracy of previous information coming from that witness, the relation between the witness and the target, and so on. It is represented by the predicate $TPImage(w, t, \varphi, \alpha, r)$ where w is the witness, t the target of w's evaluation, φ the aspect that w is evaluating about t, α the result of this evaluation and r the reliability of this $TPImage$ from the point of view of the agent.

Going up another level we find the concepts of Image and Reputation. In the *ImRep* memory, an image is represented by the predicate $Image(s, t, \varphi, \alpha, r)$ and a reputation by the predicate $Reputation(s, t, \varphi, \alpha, r)$. s and t are the source and the target agent respectively, φ the aspect that is being evaluated, α the value of that evaluation and r the reliability of that Image/Reputation.

Finally we arrive to the last memory level. All the concepts in the previous levels where linked to image and reputation. In the last level however, we find concepts represent general mental states not necessarily associated to a specific domain.

Concretely at this level the *ImRep* model considers the concepts described in section 2 (*cognitive dissonance* and *certainty*). A *cognitive dissonance* in an agent s's mind, associated to agent t and to the aspect φ is represented by the predicate $CogDis(s, t, \varphi)$. Similarly, the *certainty* of an agent s about an agent t, related to aspect φ, is represented by the predicate $Certainty(s, t, \varphi)$.

It is important to remember that this is just a conceptual organization. The memory of the agent can be just a blackboard with no internal structure as we said in section 3.1. We are not imposing a specific structure to the agent memory. That way, we make easy the integration of *ImRep* with a wide range of agent architectures.

The task of generating new predicates and maintaining the organization of the memory we have just described is performed by what we call the *detectors*, a set of specialized processes that operate on the *ImRep* memory and take care of it.

3.3 The *Detectors*

The *detectors* are one of the most important and specialized parts of the *ImRep* model. As we have said, the *detectors* are processes responsible of generating and maintaining the *ImRep* memory. They have two main tasks: (i) the inference of new and more abstract knowledge and (ii) the maintenance of the knowledge they have generated in a similar way it is done by a "'Truth maintenance system"' [4, 6, 5].

We will analyze both tasks one at a time in the following sections. However, because our final objective is to show how *ImRep* combines image and reputation we will not explain the detectors at the lower memory levels, concentrating on those detectors that actuate in the last level. At this moment the *detectors* at the lower memory levels are using the same technology used in the ReGreT system [11, 9] when they have to generate (i) an outcome from a contract and its fulfillment, (ii) an image from a set of outcomes and (iii) a reputation from witness information.

3.3.1 Generating knowledge

The *ImRep* model has a different type of detector to generate each type of predicate (except for the predicates at the lowest level that are introduced into the memory from the outside by the communication module). The fact that *detectors* are independent units allows us to design each *detector* using the most suitable technology to deal with the specific problem.

3.3.2 Cognitive dissonance detector

Talking about image and reputation, a *cognitive dissonance* appears when there is a contradiction in the agent's mind between a target's image and the reputation of that target. As we said in section 2, a cognitive dissonance generates a state of instability in the agent's mind that needs to be solved. Depending on how strong is the dissonance and the goals of the agent, this necessity becomes more or less important.

A cognitive dissonance detector uses the following rule to identify this state:

$$\begin{array}{rcl} Image(s,t,\varphi,\alpha,r) & \wedge & \\ Reputation(s,t,\varphi,\alpha',r') & \wedge & \\ IsContr(\alpha,\alpha') & \rightarrow & CogDis(s,t,\varphi) \end{array}$$

That is, if according to agent s there is a contradiction ($IsContr(\alpha,\alpha')$) between the image and the reputation of agent t related to aspect φ, then exists a cognitive dissonance related to this aspect.

How strong is the dissonance depends on three elements:

- The Image reliability (r)
- The Reputation reliability (r')
- The degree of contradiction between the image and the reputation, that is represented as $Contr(\alpha,\beta)$.

It is clear that a dissonance created by a reliable image and a (contradictory) reliable reputation creates an instability in the mind of the individual more important than a dissonance caused by a non reliable image and/or reputation.

Something similar occurs if the degree of contradiction between the image and reputation is more or less important. In the simplified case that we are considering, where the values for Image and Reputation are binary (good or Bad), this parameter has no sense. However in a real case where the values are not only Good or Bad this parameter has a great relevance. For example, it is not the same if the image suggests very-good and the reputation very-bad that if the the image suggests almost-good and the reputation slightly-bad. The cognitive dissonance in the first case is stronger than in the second.

3.3.3 Certainty

When the image and the reputation on a target is coincident we say that the perception about the target is certain. This is the opposite situation to a cognitive dissonance.

The rule to identify this state is the following:

$$\begin{array}{rcl} Image(X,Y,\varphi,\alpha,r) & \wedge & \\ Reputation(X,Y,\varphi,\alpha,r') & \rightarrow & Certainty(X,Y,\varphi) \end{array}$$

Similarly to the cognitive dissonance case, the strength of the certainty is linked to the reliability of the Image and Reputation.

3.3.4 Maintaining the *ImRep* memory

Because the agent is immersed in a dynamic environment, the perception it has about this environment changes along time. Things that were true before, now become false and vice versa. Also the mental state of the agent changes: goals that have been achieved, the appearance of new goals...

Given that, it is clear that the knowledge inferred by detectors has to be revised regularly to be sure it reflects the current situation. This revision process, that is performed by the detectors themselves, works in a similar way as a Truth Maintenance System.

Figure 2 shows a typical situation. Using witness information, detector-1 has generated a set of third party images that detector-2 has used to build a reputation. In parallel, detector-3 has inferred an Image from a set of outcomes that, at the same time, were inferred from contracts and fulfillments by detector-4. Finally detector-5, given a contradiction between the image and the reputation value, has generated a cognitive dissonance.

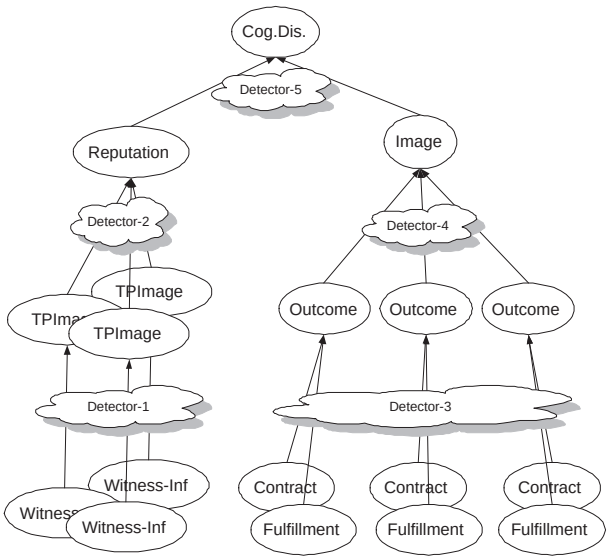

Figure 2. Mantaning the *ImRep* memory

Suppose the agent suddenly discovers that several witnesses that provided the information to build the reputation were lying. Reacting to this change, detector-1 removes the third party images associated to these witnesses that now do not have enough credibility to be taken into account. Detector-2 now realize that the number of third party images is not enough to build a reputation and the reputation

is also removed. Finally, without the reputation the cognitive dissonance has no sense so it is removed by detector-5.

Notice that the fact that the agent discovers that the witnesses were lying probably was the result of the actions taken by the agent to solve the cognitive dissonance. Once the cognitive dissonance is solved these actions, that are represented by goals in the agent's memory, are removed.

3.4 The decision making mechanism

Using *detectors*, the *ImRep* model can identify if there is a certainty or a cognitive dissonance. In the case of a cognitive dissonance it is required that the agent take different actions to solve it. We have discussed these actions in section 2 and, as we said, they are usually expensive in terms of agent resources. The decision making mechanism of the *ImRep* model is responsible of deciding if a cognitive dissonance is worth it to be solved and which are the actions to be performed in order to do that.

The main elements that the decision making mechanism uses to decide if it is worth it or not to go ahead with the actions necessary to solve a cognitive dissonance are the strength of the cognitive dissonance and, more important, the agent goals and their priority.

If the dissonance is associated to an aspect that is an essential part of a plan to achieve an important goal, the relevance of solving that dissonance has to be proportional to that importance (even if the origin of the dissonance is a weak image and/or reputation or that the level of contradiction between the image and the reputation is low). Similarly, a dissonance that according to the image and reputation should be very strong can become weak if the aspect associated to that dissonance is not relevant for the agent.

4 Conclusion

One of the main problems of current trust and reputation models is how to combine the different elements that compound trust and reputation. To undertake this problem we think it is necessary a cognitive approach.

In this article we have introduced *ImRep*, a trust and reputation model that is being designed using a cognitive approach and having in mind the problem of aggregating the elements that are used to build trust and reputation. We have made a first analysis of the interplay of Image and Reputation, and used this as a case study to illustrate the *ImRep* model.

The work presented here is at an initial stage and there are still a lot of details that need to be elaborated. Our focus was only to show the main ideas behind the *ImRep* model and the way we think it has to be solved the problem of combining the elements that are used to build trust and reputation. Future work will elaborate this ideas to obtain a fully functional computational model that can be used in different domains.

5 Acknowledgments

Jordi Sabater enjoys a Marie Curie Intra-European Fellowship (6th European framework program) contract No MEIF-CT-2003-500573.

References

[1] J. Carbo, J. Molina, and J. Davila. Trust management through fuzzy reputation. *Int. Journal in Cooperative Information Systems*, pages in–press, 2002.

[2] R. Conte and M. Paolucci. *Reputation in artificial societies: Social beliefs for social order*. Kluwer Academic Publishers, 2002.

[3] D.Huynh, N. Jennings, and N. Shadbolt. Developing an integrated trust and reputation model for open multi-agent systems. In *Proceedings of the Workshop on Trust in Agent Societies at The Third International Joint Conference on Autonomous Agents and Multi Agent Systems, New York, USA*, pages 65—74, 2004.

[4] J. Doyle. A truth maintenance system. *Artifial Intelligence*, 12(3):231–272, 1979.

[5] P. Ganderfors. *Belief Revision*. Cambridge University Press, 1992.

[6] J.P.Martins. The truth, the whole truth, and nothing but the truth: An indexed bibliography to the literature of truth maintenance systems. *AI Magazine*, 11(5):7–25, 1991.

[7] S. Marsh. *Formalising Trust as a Computational Concept*. PhD thesis, Department of Mathematics and Computer Science, University of Stirling, 1994.

[8] M. Paolucci. False reputation in social control. *Advances in Complex Systems*, 3(1-4):39–51, 2000.

[9] J. Sabater. *Trust and reputation for agent societies*. PhD thesis, Universitat Autonoma de Barcelona (UAB), 2003.

[10] J. Sabater. Evaluating the regret system. *The Journal of Applied Artificial Intelligence*, 18(9-10):in–press, 2004.

[11] J. Sabater and C. Sierra. Reputation and social network analysis in multi-agent systems. In *Proceedings of the first international joint conference on autonomous agents and multiagent systems (AAMAS-02), Bologna, Italy*, pages 475—482, 2002.

[12] J. Skowronski and D. Carlston. Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, 105:131–142, 1989.