A Bayesian Computational Model for Trust on Information Sources

Alessandro Sapienza and Rino Falcone
Institute of Cognitive Sciences and Technologies, ISTC – CNR,
Rome, Italy
{alessandro.sapienza, rino.falcone}@istc.cnr.it

Abstract— **In this work we want to provide a tool for handling information coming from different information sources. In fact the real world we often have to deal with different sources asserting different things and, in order to decide, it is necessary to consider properly each of them trying to put this information together. According to us, a good way to do it is exploiting the concept of trust. In fact using it as a valve, it is possible to give a different weight to what the source is reporting. Plus we decide to implement this trust model as generic as possible. In this way, the model can be used in different context and within different practical applications.**

After presenting the theoretical and the computational model, we also show a practical example of how to use it, to let the reader better understand the overall workflow.

Keywords—trust; cognitive model; bayesian theory

I. INTRODUCTION

In the world we often have to deal with different information coming from different information sources. Though having a lot of sources can be very useful, on the other hand, trying to put together information coming from different information sources can be an uneasy task. It is necessary to have strategies to do it, especially in presence of critical situation, when there are temporal limits to make decision and a wrong choice can lead to an economical loss or even to risk life.
As said, the possibility of integrating sources on different scopes can be very useful in order to make a well-informed decision.
Integrating these sources becomes essential, but at the same time it is necessary to identify and take into account their trustworthiness.

In our perspective [3][4] trust in information sources is just a kind of social trust, preserving all its prototypical properties and dimensions; just adding new important features and dynamics. In particular, also the trust in information sources [6] can just be an evaluation, judgment and feeling, or be a decision to rely on, and act of believing in and to the trustee (Y) and rely on it. Also this trust and has two main dimensions: the ascribed competence versus the ascribed willingness (intentions, persistence, reliability, honesty, sincerity, etc.).
Moreover this form of trust is not empty, but it possesses a more or less specified argument: the trustor X can not just trust Y, as trust is for/about something, it has a specific object: what X expects from Y; Y's service, action, provided well. And it is also context-dependent: in a given situation; with internal or external causal attribution in case

Then, according to our view [3] trusting an information source (S) means to use a cognitive model based on the dimensions of competence and motivation of the source. These competence and motivation evaluations can derive from different reasons, basically:

- Our previous *direct experience* with S on that specific kind of information content.
- *Recommendations* (other individuals Z reporting their direct experience and evaluation about S) or *Reputation* (the shared general opinion of others about S) on that specific information content [5][11][15][16][19].
- *Categorization* of S (it is assumed that a source can be categorized and that it is known this category), exploiting inference and reasoning (analogy, inheritance, etc.): on this basis it is possible to establish the competence/reliability of S on that specific information content [1][2][7][8]. In past works, we showed that exploiting categories for trust evaluations can represent a significant advantage [9][10].

Considering information's output, it can be a true/false one (the source can just assert or deny the belief P) or there can be multiple outcomes. As this is a general model, we suppose that there can be different outcomes. For instance, the weather is not just good or bad, but can assume multiple values (critical, sunny, cloudy etc.).

II. The Bayesian Choice

There are many ways to computationally realize a decision making process and quite all of them provide good results.
Dealing with uncertain situations, one can use the uncertainty theory [12], a mathematical approach specifically created to evaluate belief degree in cases in which there is no data.
Another possible way is to use fuzzy logic [17]. This technique has several vantages like:

1. It is flexible and easy to use;
2. It don't need precise data;
3. It can deal with non linear functions;
4. It is able to shape human way of think and express, as it can model concept that are more complex than a Boolean but not so precise like a real number.

Maybe the most used approach is the probabilistic one, which exploits the Bayesian theory, in particular probability distribution.
One of the advantages of using Bayesian theory is that it implies a sequential process: every time that new evidence occurs it can be processed individually and then aggregated to global evidence. This property is really useful as it allows a trustor to elaborate its information in a moment and update it whenever it gets other evidence.

Given the context of information sources, we believe that this last option is the choice that best suits with the problem. In fact there is a fixed number of known possibilities to model and the trustor can collect information from its sources individually and then aggregate them in different moment.
Plus, the scientific literature confirms its utility in the context of trust evaluation[13][14][18].

III. The Computational Model

In the proposed model each information source S is represented by a trust degree called $TrustOnS$, with 0≤ $TrustOnS$ ≤1, plus a bayesian probability distribution PDF that represents the information reported by S.

To the aim of granting a better flexibility, the PDF is modeled as a continuous distribution (actually it is divided into several intervals and it is continuous in each interval). In fact if the event domain is continuous it is better to use a continuous PDF; if it happens to be discrete it is still possible to use a continuous PDF. It is also possible to specify what and how much outcomes the model has to use, depending on the specific context. In the end of the paper we will show a working example in which we take into account five different outcomes, then the PDF will be divided accordingly.

The model we created starts from a preliminary evaluation of the source trustworthiness: how much reliable is a source S concerning a specific information's category?
Then after evaluating it, we consider what the source is reporting - the PDF. We use the trust evaluation to understand how much the specific information should be considered, with respect to the global information.

This process can be done in presence of a single or multiple sources, as each time we perform an aggregation of each contribute to the global evidence.
A strong point of this model is that it is sequential, so it can be updated when new information comes.

A. Source's Evaluation

The first part of the model concerns the source's evaluation. According to us, there are two level of evaluation. Initially, we produce an a priori trust, which represent how much I believe that S is good with this specific kind of information.
After that, we compute a more sophisticated analysis taking into account other parameters.

Let's first start from the a priori source's evaluation – $SEvaluation$. **This is the trustor's trust about P just depending on the its judgment of the S's competence and willingness** as derived from the composition of the three factors (direct experience, recommendation/reputation, and categorization), in practice the S's credibility about P on view of the trustor.
Recalling that a trust evaluation for a cognitive agent is based on the two aspects of competence and willingness, we state that these values can be obtained using three different dimensions:

1. **Direct experience** with S (how S performed in the past interactions) on that specific information content;
2. **Recommendations** (other individuals Z reporting their direct experience and evaluation about S) or **Reputation** (the shared general opinion of others about S) on that specific information content;
3. **Categorization** of S.

The two faces of S's trustworthiness (competence and willingness) are relatively independent; however, for sake of simplicity, we will unify them into a unique quantitative parameter, by combining competence and reliability.

Computationally, the past experience (PE), reputation/recommendation (REP) and categories (CAT) parameters are defined here as real values in the interval [0,1]. To compute S's evaluation we make a weighted mean of them:

$$SEvaluation = w1 * PE + w2 * CAT + w3 * REP$$

The trustor, considering both its personality and the context in which it is, determines the weight w1, w2 and w3 empirically.

B. Certainty and Identity

Computing the general trust on the Source concerning P is a good starting point. However it is not enough. In fact, while this value represents an a priori evaluation of how much a source S is trustworthy, there are other two factors that can influence a trust evaluation.

The first one is the **S's degree of certainty about P ($Certainty$)**. The information sources not only give the information but also their certainty about this information. The same information can be reported with different degree of confidence ("I am sure about it", "I suppose that", "it is possible that" and so on).

Of course we are interested in modeling this certainty, but we have to consider that through the trustor's point of view (it subjectively estimates this parameter). It is defined as a real value in range [0,1].

The second dimension represents **the trustor's degree of trust that P derives from S ($Identity$)**: the trust we have that the information under analysis derives from that specific source; it is defined as a real value in range [0,1]. This parameter has a twofold meaning:

1. For instance, considering the human communication I can be more or less sure that the specific information under analysis has been reported by the source S. It is a problem of memory, do I recall properly?
2. In the web context the communication's dynamics changes. I will probably receive the information by someone hiding beyond a computer. How may I be sure about it's identity? Can I trust that S is really who is saying to be? This is a very complex issue and its solution has not been completely provided by computer scientist.

The source Evaluation is softened by the Certainty and the Identity parameters, since we considered them as two multiplicative parameters. The output of this operation is the actual trust that the trustor has on S:

$$TrustOnS = SEvaluation * Identity * Certainty$$

C. PDF: the reported information

With the PDF (Probability Distribution Function) we represent the probability distribution that the source reports concerning the belief P.

Given a fixed number of outcomes, which depends on the nature of the information and on the accuracy of the source in reporting the information, with the PDF a source S reports how much it subjectively believes possible each single outcome.

Of course the source can assert that just one of them is possible (100%) or it can divide the probability among them.

The picture 1 shows an example of what we mean with the term PDF. It is divided in slots, each one representing a possible outcome.

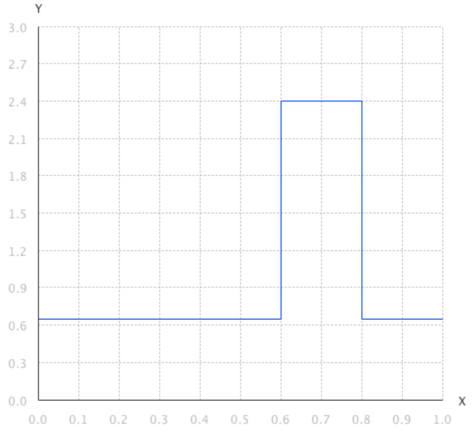

Figure 1: An example of a PDF

It is not possible to consider the PDF as it is. The idea is that if I think I am exploiting a reliable source, than it is good to take into account what it is saying. But if I suppose that the source is unreliable, even if it is not competent or because there is a possibility it wants to deceive me, then I need to be cautious.

Here we propose an algorithm to deal with this problem, combining the trust evaluation with what the source is reporting. In other words, we exploit the $TrustOnS$ value to smooth the PDF. The output of this process is what we call the Smoothed PDF (SPDF).

Recalling that the PDF is divided into segments, this is the formula used for transforming each segments:

$$Segment_i = 1 + (Segment_i - 1) * TrustOnS$$

If $Segment_i > 1$ it will be lowered until 1. On the contrary, if $Segment_i < 1$it will tend to increase to the value 1.

We will have that:

- The greater $TrustOnS$ is, the more similar the SPDF will be to the PDF; in particular if $TrustOnS$ =1 => SPDF =PDF;
- The lesser it is, the more the SPDF will be flatten; in particular if $TrustOnS$ =0 => SPDF is an uniform distribution with value 1.

The idea is that we trust on what S says proportionally to how much we trust S. In words, the more we trust S, the more we tend to take into consideration what it says; the less we trust S, the more we tend to ignore its informative contribution.

The picture 2 resumes the model until this point.

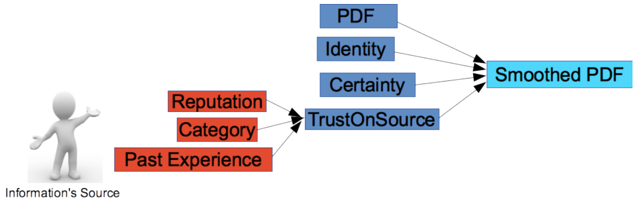

Figure 2: A scheme of the computational model until the SPDF

D. The effect of each source/evidence on the Global PDF

We define GPDF (Global PDF) the evidence that an agent owns concerning a belief P. At the beginning, if the trustor does not possess any evidence about the belief P, the GPDF is flat, as it is a uniform distribution. Otherwise it has a specific shape the models the specific internal belief of the trustor.

Each information source provides evidence about P, modifying then the GPDF owned by the trustor. Once estimated the SPDFs for each information source, there will be a process of aggregation between the GPDF and the SPDFs. Each source actually represents a new evidence E about a belief P. Then to the purpose of the aggregation process it is possible to use the classical Bayesian logic, recursively on each source:

$$f(P|E) = \frac{f(E|P) * f(P)}{f(E)}$$

where:
f(P|E) = GPDF (the new one)
f(E|P) = SPDF;
f(P) = GPDF (the old one)

In this case f(E) is a normalization factor, given by the formula:

$$f(E) = \int f(E|P) * f(P)\, dP$$

In words the new GPDF, that is the global evidence that an agent has about P, is computed as the product of the old GPDF and the SPDF, that is the new contribute reported by S.
As we need to ensure that GPDF is still a probability distribution function, it is necessary to scale down it[1]. This is ensured by the normalization factor f(E).

The picture 3 represents the whole model for managing trust on information sources

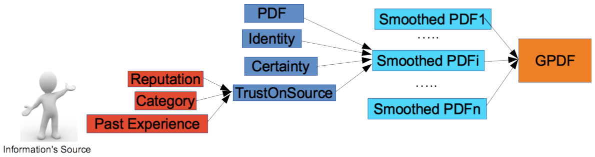

Figure 3: A scheme of the computational model until the GPDF

Exploiting the GPDF, the trust is able to understand what is the outcome O_i that is more likely to happen.

E. Handling uncertainty

Dealing with information, a critical point is how to handle uncertainty.
The point is that considering uncertainty on information is correct, but it is a too limitative approach. In fact uncertainty comes up at different levels and has to be taken into account when deciding.
Actually, in this model we handle it in three different ways.

The first one is the **uncertainty on the source**. This is given by the source evaluation $SEvaluation$.
The second level is represented by **uncertainty on communication**. This is handled by the two parameters Certainty and Identity: how much I'm sure about the identity of the source? How much certainty does the source express in reporting the information (according to the trustor)?
The last level is the **uncertainty on the reported information** (PDF). This is managed just by the intrinsic nature of the PDF. In fact what happens here is that the source express its certainty/uncertainty through the outcomes' distributions.

In practice, we take into account uncertainty in all the process, until the end, in order to produce a proper prediction.

IV. A Workflow's Example

In this section we want to provide a working example of how to use the model. As the trust computation is quite simple and intuitive, below we will directly use the TrustOnS parameter, together with the corresponding PDF.
Moreover, we will represent PDFs as a list of five values, with the following formalism:

$$PFD_{Si} = [x_{i1}\; x_{i2}\; x_{i3}\; x_{i4}\; x_{i5}]$$

in which $x_{i1}\; x_{i2}\; x_{i3}\; x_{i4}\; x_{i5}$[2] are the values of the PDF for the source S_i in the corresponding segment.

Suppose that an agent has to understand what kind of weather there will be the following day. It starts collecting forecast from its information sources. The possible outcomes are five: {sunny day, cloudy day, light rain, heavy rain, critical rain}.

Let's suppose that Source S1 has a $TrustOnS_{S1}$=1 (the maximal value) and that it is asserting PDF_{S1} = [0.5 0.5 0.5 3 0.5], so it mainly suppose that there will be heavy rain.
The visual representation of PDF_{S1} is provided by figure 4.

[1] To be a PDF, it is necessary that the area subtended by it is equal to 1.

[2] Note that, from how the PDF has been defined, these parameters are non-negative real numbers, with the peculiarity that their sum is equal to 5.

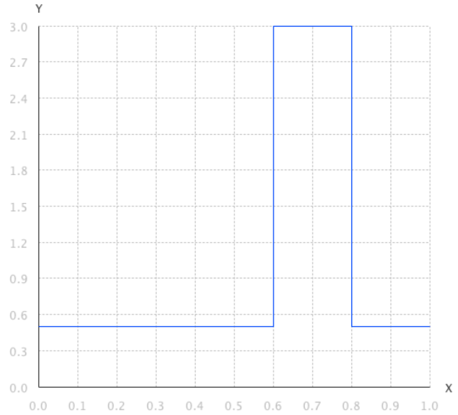

Figure 4: The representation of PDF$_{S1}$ in the example

As the trustor has the maximal trust on S1, PDF$_{S1}$ and SPDF$_{S1}$ will be the same. Plus, as this the first evidence on P, even the GPDF is equal to PDF$_{S1}$.

Let then see what happens to S2, asserting the same of S1, but with a TrustOnSource$_{S2}$ of 0.7. The PDF$_{S2}$ is the same of PDF$_{S1}$, but the SPDF$_{S2}$ is different, as showed by figure 5:

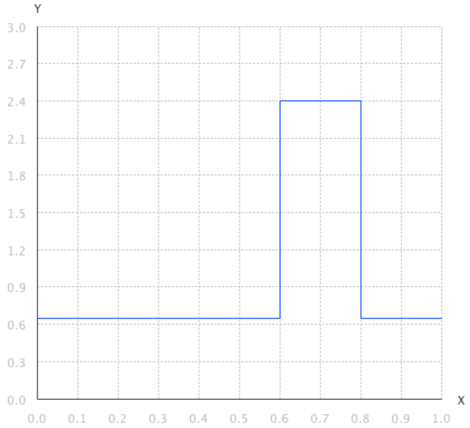

Figure 5: The representation of SPDF$_{S2}$ in the example

The PDF$_{S2}$ has been smoothed, so that values grater than 1 has been decreased and values smaller than one has been increased.

Let's then see what happens to the GPDF:

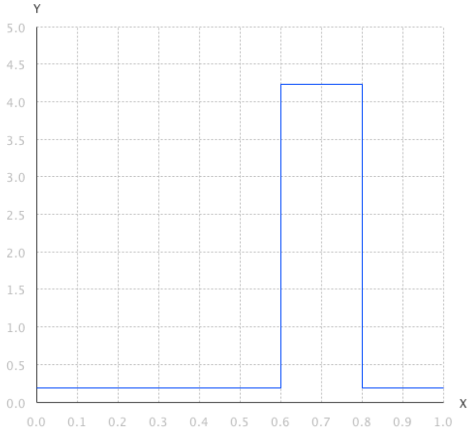

Figure 6: The representation of GPDF in the example with the contribute of S$_1$ and S$_2$.

As showed by figure 6, Thanks to the fact that the sources, even if with two different trust degrees, are asserting the same things, there is a reinforcement of evidence in segment 4 of the GPDF.
This is a peculiarity that we shaped in our previous models and that persist in this one as a consequence of the Bayes theorem.

Let's than see what happen in presence of a third source S3, with TrustOsSource$_{S3}$ = 0.3 and PDF$_{S3}$ = [0.3 3.8 0.3 0.3 0.3]. This source is reporting a cloudy day forecast. Its SPDF will be:

The final result is showed by figure 7:

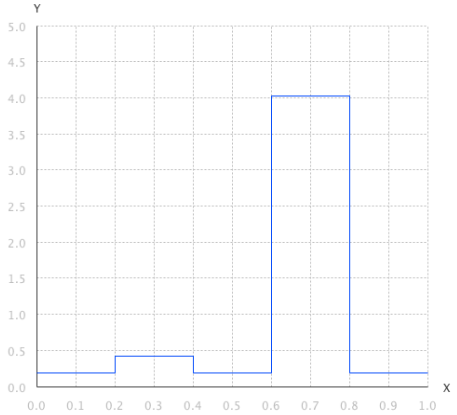

Figure7: The final representation of GPDS in the example

The new GPDF is quite the same of the previous one. This is due to the fact that, although S3 is strongly disagreeing with S1 and S2, it has a low level of trust. Then it will lightly affect what the trustor thinks.
In the end the trustor can assert that there will be heavy rain the next day.

V. Conclusion

The aim of this work was that of realizing a theoretical and computational model for dealing with information sources.

This is in fact an uneasy task and there can be critical situations in which agents have to face sources asserting different things.

We decided to realize a model as generic as possible. Doing so, the model does not depend on a specific context and it can be applied on different practical context.

The basic idea is that using trust on information sources is a promising way to face the problem. Then, from a theoretical point of view, we analyzed all the possible cognitive variables that can affect trust on an information source.

After analyzing the various ways to represent information, we decided to exploit Bayesian theory. Then we showed how to apply the trust evaluation on the information layers in order to properly take into account information.

Finally, we proposed a practical problem – the one of weather forecast – and we showed how to apply the model in order to get a solution.

Acknowledgment

This work is partially supported by the project CLARA—CLoud plAtform and smart underground imaging for natural Risk Assessment, funded by the Italian Ministry of Education, University and Research (MIUR-PON).

References

[1] Burnett, C., Norman, T., and Sycara, K. 2010. Bootstrapping trust evaluations through stereotypes. In Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'10). 241-248.

[2] Burnett, C., Norman, T., and Sycara, K. (2013) Stereotypical trust and bias in dynamic multiagent systems. ACM Transactions on Intelligent Systems and Technology (TIST), 4(2):26, 2013.

[3] Castelfranchi, C., Falcone R., Pezzulo, (2003) Trust in Information Sources as a Source for Trust: A Fuzzy Approach, Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-03) Melburne (Australia), 14-18 July, ACM Press, pp.89-96.

[4] Castelfranchi C., Falcone R., Trust Theory: A Socio-Cognitive and Computational Model, John Wiley and Sons, April 2010.

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

[6] Demolombe R., (1999), To trust information sources: A proposal for a modal logic frame- work. In Castelfranchi C., Tan Y.H. (Eds), Trust and Deception in Virtual Societies. Kluwer, Dordrecht.

[7] Falcone R, Castelfranchi C, (2008) Generalizing Trust: Inferencing Trustworthiness from Categories. In Proceedings, pp. 65 - 80. R. Falcone, S. K. Barber, J. Sabater-Mir, M. P. Singh (eds.). Lecture Notes in Artificial Intelligence, vol. 5396. Springer, 2008

[8] Falcone R., Piunti, M., Venanzi, M., Castelfranchi C., (2013), From Manifesta to Krypta: The Relevance of Categories for Trusting Others, in R. Falcone and M. Singh (Eds.) Trust in Multiagent Systems, ACM Transaction on Intelligent Systems and Technology, Volume 4 Issue 2, March 2013

[9] Falcone, R., Sapienza, A., & Castelfranchi, C. (2015, July). Recommendation of categories in an agents world: The role of (not) local communicative environments. In *Privacy, Security and Trust (PST), 2015 13th Annual Conference on* (pp. 7-13). IEEE.

[10] Falcone, R., Sapienza, A., & Castelfranchi, C. (2015). The relevance of categories for trusting information sources. *ACM Transactions on Internet Technology (TOIT)*, *15*(4), 13.

[11] S. Jiang, J. Zhang, and Y.S. Ong. An evolutionary model for constructing robust trust networks. In Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), 2013.

[12] B. Liu, Uncertainty theory 5th Edition, Springer 2014.

[13] Melaye, D., & Demazeau, Y. (2005). Bayesian dynamic trust model. In Multi-agent systems and applications IV (pp. 480-489). Springer Berlin Heidelberg.

[14] Quercia, D., Hailes, S., & Capra, L. (2006). B-trust: Bayesian trust framework for pervasive computing. In Trust management (pp. 298-312). Springer Berlin Heidelberg.

[15] Sabater-Mir, J. 2003. Trust and reputation for agent societies. Ph.D. thesis, Universitat Autonoma de Barcelona

[16] Sabater-Mir J., Sierra C., (2001), Regret: a reputation model for gregarious societies. In 4th Workshop on Deception and Fraud in Agent Societies (pp. 61-70). Montreal, Canada.

[17] Sapienza, A., Falcone, R., & Castelfranchi, C. Trust on Information Sources: A theoretical and computation approach, in proceedings of WOA 2014, ceur-ws, vol 1260, paper 12.

[18] Wang, Y., & Vassileva, J. (2003, October). Bayesian network-based trust model. In Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference on (pp. 372-378). IEEE.

[19] Yolum, P. and Singh, M. P. 2003. Emergent properties of referral systems. In Proceedings of the 2nd International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS'03).