Goal Generation from Beliefs Based on Trust and Distrust

(Extended Abstract)

Célia da Costa Pereira and Andrea G. B. Tettamanzi
Università degli Studi di Milano
Dipartimento di Tecnologie dell'Informazione
Via Bramante 65, I-26013 Crema (CR), Italy
pereira@dti.unimi.it, andrea.tettamanzi@unimi.it

ABSTRACT

We propose three alternative belief change operators to be used in a goal generation and adoption framework. The originality of these operators is that the trustworthiness of a source depends not only on the degree of trust but also on an *independent* degree of distrust. Explicitly taking distrust into account allows us to mark a clear difference between the distinct notions of *negative* trust and *insufficient* trust. More precisely, it is possible, unlike in approaches where only trust is accounted for, to "weigh" differently information from *helpful*, *malicious*, *unknown*, or *neutral* sources.

Categories and Subject Descriptors

I.2.3 [**Artificial Intelligence**]: Deduction and Theorem Proving—*Nonmonotonic reasoning and belief revision*

General Terms

Theory

Keywords

Beliefs, desires and goals, fuzzy logic

1. INTRODUCTION AND MOTIVATION

Recently, a belief change operator for goal generation has been proposed [4]. There, it is supposed that the sources of information are not merely divided into the trustworthy and the malicious in a clear-cut way, and it is shown that the set of goals to be generated by an agent may then depend on the extent to which the agent trusts the information sources.

The main lack in that approach is the way the concept of *distrust* is implicitly considered, that is, as the complement of *trust*. Thus doing, trusting a source to a degree $\tau \in [0, 1]$ means distrusting that source to degree $1 - \tau$. However, things are not always so simple. Trust and distrust may derive from different kinds of information and, therefore, can coexist without however being complementary. In many situations, like in politics for example, two persons or two nations can trust each other enough to cooperate, but at the same time maintain a certain degree of distrust. Besides, the fact that someone does not completely trust a source does not always mean that person has a reason to distrust the source. It may only mean that (s)he does not have any reason or enough evidence to lead her/him to completely trust the source.

We propose a way to bridge this gap by concentrating on the influence of new information in the agent's beliefs and by considering explicitly not only the trust degree in the source but also the distrust degree. The trustworthiness of a source is represented as a pair (trust, distrust), and intuitionistic fuzzy logic [1] is used to represent the uncertainty, introduced by the presence of distrust, on the trust values. On that basis, we propose three belief change operators which mimic the agent attitudes towards information coming from trusted, malicious, neutral (trusted and malicious at same extent) or unknown sources.

The paper is organized as follows: Section 2 briefly motivates and discusses a bipolar view of trust and distrust; Section 3 proposes three possible generalizations of a trust-based fuzzy belief change operator for dealing with explicitly given trust and distrust; Section 4 briefly outlines the impact of belief change on desire and goal generation and change. Finally, Section 5 concludes the paper.

2. TRUST AND DISTRUST

Our aim is not to compute degrees of trust and distrust of sources; we are just interested in how these degrees influence the agent's beliefs and, as a consequence, the choice of which set of goals, among the possible ones, the agent will generate/adopt. Approaches to the problem of assigning degrees of trust (and distrust) to information sources can be found, for example, in previous work by Castelfranchi and colleagues [3], McKnight and Chervany [9], and by De Cock and da Silva [6].

We propose to define the *trustworthiness* score of a source as a pair $(\tau_s, \delta_s) \in [0, 1] \times [0, 1]$, where τ_s is the degree to which source s is trusted and δ_s is the degree to which source s is distrusted, with $\tau_s + \delta_s \leq 1$. We also define the *hesitation degree*, $h = 1 - \tau - \delta$, which corresponds to the length of the interval of the possible values of trust $[\tau, 1 - \delta]$, [7].

The agent perception of a source depends on the extent of its reasons to believe (or not) in the source. A *helpful* source is a source for which the reasons to believe are stronger than the reasons to reject its information. A *malicious* source is a source for which the reasons to reject its information are stronger than the reasons to believe it. An *unknown* source is a source which never provided information to the agent before. Finally, a *neutral* source is a source which provided as much true information as false.

3. BELIEF CHANGE

The formalism adopted here is that used in [4]. In such a formalism an agent's mental state $\mathcal{S}$ is completely described by a triple $\mathcal{S} = \langle \mathcal{B}, \mathcal{R}_J, \mathcal{J} \rangle$, where (i) $\mathcal{B}$ is a fuzzy interpretation on atomic propositons representing the agent's beliefs; (ii) $\mathcal{R}_J$ contains the

Cite as: Goal Generation from Beliefs Based on Trust and Distrust, (Extended Abstract), Célia da Costa Pereira, Andrea G. B. Tettamanzi, *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009),* Decker, Sichman, Sierra and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary, pp. 1127–1128

rules which generate desires from beliefs and other desires (subdesires); and (iii) $\mathcal{J}$ is a fuzzy set of literals representing the agent's desires.

Here, we propose three possible extensions for the belief change operator $*$ proposed in [4], to deal with bipolar trust and distrust degrees.

3.1 Open-Minded Belief Change Operator

This operator represents the changes in the beliefs of an agent which is both optimistic and does not perceive malicious purposes from neutral sources. The proposed operator provides a formal representation of how an agent which gives the benefit of the doubt to the sources could change its beliefs when new information is received.

Let ϕ be a formula, and $\mathcal{B}$ be the agent's beliefs. An *open-minded operator* $*_m$ can be defined as

$$\mathcal{B} *_m \frac{(\tau, \delta)}{\phi} = \mathcal{B} * \frac{\tau + (h/2)}{\phi}. \quad (1)$$

where h is the hesitation degree of (τ, δ) (i.e., $h = 1 - \tau - \delta$).

As we can see, such an agent chooses a degree of trust which is proportional to the degree of hesitation. The greater the hesitation, the higher the adopted trust degre.

3.2 Wary Belief Change Operator

Here, we present a belief change operator, $*_w$, which illustrates the attitude of a prudent agent which does not give the benefit of the doubt to unknown sources and perceives information coming from malicious sources as false (to some extent, depending on the degree of distrust). Operator $*_w$ can be defined as

$$\mathcal{B} *_w \frac{(\tau, \delta)}{\phi} = \begin{cases} \mathcal{B} * \frac{\tau - \delta}{\phi} & \text{if } \tau > \delta \text{ and } h \neq 0; \\ \mathcal{B} * \frac{\delta - \tau}{\neg\phi} & \text{if } \delta > \tau \text{ and } h \neq 0; \\ \mathcal{B} * \frac{0}{\phi} & \text{if } \tau = \delta. \end{cases} \quad (2)$$

The last condition, $\tau = \delta$, could be integrated into one of the previous two by considering $\tau \geq \delta$ or $\delta \geq \tau$, but, for the sake of clarity, we have preferred to express it separately.

3.3 Content-Based Belief Change Operator

Neither of the previous two proposed operator extensions attempts to consider the content of incoming information as well as the agent's competence in judging its truth.

Experiments on human trust and distrust of information, like those carried out by McGuinness and Leggatt [8], have shown that if what the source is saying finds confirmations within the beliefs one already entertains, the source is rather regarded as trustworthy because of that. If, on the contrary, what the source is saying goes against one's previous beliefs, the tendency is to discredit the source. The basic rationale for this behaviour appears to be that people trust themselves, if anybody.

The third proposed extension of the belief change operator, $*_c$, intends to mimic, or model, this type of behavior. Operator $*_c$ can be defined as

$$\mathcal{B} *_c \frac{(\tau, \delta)}{\phi} = \mathcal{B} * \frac{\tau + h \cdot \mathcal{B}(\phi)}{\phi}, \quad (3)$$

where $h = 1 - \tau - \delta$.

We can notice that by applying operator $*_c$, the only cases in which the information content does not influence the agent's beliefs is when there is no hesitation, i.e., $h = 0$.

4. DESIRE AND GOAL CHANGE

Belief change, according to [2], is the only way to modify the agent goals from the outside (external reasons).

To account for the changes in the desire set caused by belief change in the case of the $*$ operator, and, as a consequence, also in case of the three proposed operators $*_m$, $*_w$, and $*_c$, one has to recursively: (i) calculate for each rule $R \in \mathcal{R}_\mathcal{J}$ its new activation degree by considering the updated agent's beliefs, $\mathcal{B}'$, and (ii) update the justification degree of all desires in its right-hand side, until a fixed point is reached [5, 4].

In general, given a fuzzy set of desires $\mathcal{J}$, there may be more than one possible goal set. However, a rational agent, for practical reasons, may need to elect one precise set of goals to pursue, which depends on its mental state.

The choice of one goal set over the others may be based on a preference relation on desire sets, but exactly how that is achieved falls out of the scope of this paper.

5. CONCLUSION

The issue of how to deal with independently and explicitly given trust and distrust degrees within the context of goal generation has been approached by generalizing a recently proposed fuzzy belief change operator.

The three proposed alternative extensions have different scopes. The open-minded operator makes sense in a collaborative environment, where all sources of information intend to be helpful, except that, perhaps, some of them may lack the knowledge needed to help. The wary operator is well suited to contexts where competition is the main theme and the agents are utility-driven participants in a zero-sum game, where a gain for an agent is a loss for its counterparts. The content-based operator is aimed at mimicking the usual way people change their beliefs.

6. REFERENCES

[1] K. T. Atanassov. Intuitionistic fuzzy sets. *Fuzzy Sets Syst.*, 20(1):87–96, 1986.

[2] C. Castelfranchi. Reasons: Belief support and goal dynamics. *Mathware and Soft Computing*, 3:233–247, 1996.

[3] C. Castelfranchi, R. Falcone, and G. Pezzulo. Trust in information sources as a source for trust: a fuzzy approach. In *Proceedings of AAMAS'03*, pages 89–96, 2003.

[4] C. da Costa Pereira and A. Tettamanzi. Goal generation and adoption from partially trusted beliefs. In *Proceedings of ECAI 2008*, pages 453–457. IOS Press, 2008.

[5] C. da Costa Pereira and A. Tettamanzi. Goal generation with relevant and trusted beliefs. In *Proceedings of AAMAS'08*, pages 397–404. IFAAMAS, 2008.

[6] M. De Cock and P. Pinheiro da Silva. A many valued representation and propagation of trust and distrust. In *WILF'05*, pages 114–120, 2005.

[7] G. Deschrijver and E. E. Kerre. On the relationship between some extensions of fuzzy set theory. *Fuzzy Sets Syst.*, 133(2):227–235, 2003.

[8] B. McGuinness and A. Leggatt. Information trust and distrust in a sensemaking task. In *Command and Control Research and Technology Symposium*, 2006.

[9] D. H. McKnight and N. L. Chervany. Trust and distrust definitions: One bite at a time. In *Proceedings of the workshop on Deception, Fraud, and Trust in Agent Societies*, pages 27–54. Springer-Verlag, 2001.