Available online at www.sciencedirect.com
ScienceDirect
Cognitive Systems Research 14 (2012) 60–83
Cognitive Systems RESEARCH
www.elsevier.com/locate/cogsys

Cognitive and neural modeling of dynamics of trust in competitive trustees ☆

Action editors: Minho Lee, Wlodzislaw Duch

Mark Hoogendoorn, S. Waqar Jaffry *, Jan Treur

Vrije Universiteit Amsterdam, Department of Artificial Intelligence, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands

Received 1 April 2010; accepted 12 December 2010
Available online 15 January 2011

Abstract

Trust dynamics can be modeled in relation to experiences. In this paper two models to represent human trust dynamics are introduced, namely a model on a cognitive level and a neural model. These models include a number of parameters, providing the possibility to express certain relations between trustees. The behavior of each of the models is further analyzed by means of simulation experiments and formal verification techniques. Thereafter, both models have been compared to see whether they can produce patterns that are comparable. As each of the models has its own specific set of parameters, with values that depend on the type of person modeled, such a comparison is non-trivial. To address this, a special comparison approach is introduced, based on mutual mirroring of the models in each other. More specifically, for a given parameter values set for one model, by an automated parameter estimation procedure the most optimal values for the parameter values of the other model are determined in order to show the same behavior. Roughly spoken the results are that the models can mirror each other up to an accuracy of around 90%.

Keywords: Trust dynamics; Cognitive; Neural; Comparison; Parameter tuning

1. Introduction

Nowadays, more and more ambient systems are being deployed to support humans in an effective way (Aarts, Collier, van Loenen, & Ruyter, 2003; Aarts, Harwig, & Schuurmans, 2001; Riva, Vatalaro, Davide, & Alcañiz, 2005). An example of such an ambient system is a personal agent that monitors the behavior of a human executing certain complex tasks, and gives dedicated support for this. Such support may include advising the use of a particular information source, system or agent to enable proper execution of the task, or even involving such a system or agent pro-actively. In order for these personal agents to be accepted and useful, the personal agent should be well aware of the habits and preferences of the human. If a human for example for good reasons dislikes using a particular system or agent, and there are several alternatives available that are more preferred, the personal agent would

☆ Parts of this paper are based on work presented as: Hoogendoorn, M., Jaffry, S.W., and Treur, J., Modeling Dynamics of Relative Trust of Competitive Information Agents. In: Klusch, M., Pechoucek, M., Polleres, A. (eds.), Proc. of the 12th International Workshop on Cooperative Information Agents, (CIA'08), Prague, (Czech Republic), September 10th–12th, 2008. Lecture Notes in Artificial Intelligence, vol. 5180, Springer Verlag, 2008, pp. 55–70; Jaffry, S.W., and Treur, J., Comparing a Cognitive and a Neural Model for Relative Trust Dynamics. In: Proceedings of 16th International Conference on Neural Information Processing. (ICONIP'09), Bangkok, (Thailand), December 1st–5th, 2009. Lecture Notes in Computer Science, vol. 5863 Part I, Springer Verlag, 2009, pp. 72–83; Hoogendoorn, M., Jaffry, S.W., and Treur, J., Modelling Trust Dynamics from a Neurological Perspective. In: Proceedings of the 2nd International Conference on Cognitive Neurodynamics. (ICCN'09), Hangzhou, (China), November 15th–19th, 2009.

* Corresponding author.
E-mail addresses: mhoogen@few.vu.nl (M. Hoogendoorn), swjaffry@few.vu.nl (S.W. Jaffry), treur@few.vu.nl (J. Treur).

1389-0417/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogsys.2010.12.011

not be supporting effectively if it would advise, or even proactively initiate, the disliked option.

An aspect that plays a crucial role in giving such tailored advice is to represent the trust levels the human has for certain options. Knowing these trust values allows the personal assistant to reason about these levels, and give the best possible support that is in accordance with the habits and preferences of the human. Since there would be no problem in case there is only one way of supporting the human, the problem of selecting the right support method only occurs in case of substitutable options. Therefore, a notion of relative trust in these options seems more realistic than having a separate independent trust value for each of these options. For instance, if three systems or agents can contribute X, and two of them perform bad, whereas the third performs pretty bad as well, but somewhat better than the others, trust in that third option may still be relatively high since in the context of the other options it is the best alternative. The existing trust models do however not explicitly handle such relative trust notions (see e.g. Falcone & Castelfranchi, 2004; Jonker & Treur, 1999; Marx & Treur, 2001).

In this paper, a cognitive and a neural model are presented that address the dynamics of trust, including the aforementioned notion of relative trust and particular other personality characteristics. Both models are evaluated using simulation experiments and formal verification techniques.

The first model, representing trust on a cognitive level, takes into account two main functional properties of trust states, which define the *causal or functional role* of a trust state as cognitive state, as put forward in (Jonker & Treur, 2003):

(1) A trust state results from accumulation of experiences over time
(2) Trust states affect decision making by choosing more trusted options above less trusted options

The second model of trust dynamics is based on neurological principles. In this model, theories on the interaction between affective and cognitive states (see e.g., Eich, Kihlstrom, Bower, Forgas, & Niedenthal, 2000; Forgas, Goldenberg, & Unkelbach, 2009; Forgas, Laham, & Vargas, 2005; Niedenthal, 2007; Schooler & Eich, 2000; Winkielman, Niedenthal, & Oberman, 2009) are modeled on a neurological level as well by using theories on the embodiment of emotions as described, for example, in (Damasio, 1994, 1996, 1999, 2003; Winkielman et al., 2009). Based on these the model describes how trust dynamics relates to experiences with (external) sources, both from a cognitive and affective perspective. More specifically, in accordance with, for example (Damasio, 1999, 2003), for feeling the emotion associated to a mental state, a converging *recursive body loop* is assumed. In addition, based on *Hebbian learning* (cf. Bi & Poo, 2001; Hebb, 1949; Gerstner & Kistler, 2002) for the strength of the connections to the emotional responses, an adaptation process is introduced, inspired by the Somatic Marker Hypothesis (Damasio, 1994, 1996).

Being described on a different level, each of the models includes specific set of parameters for cognitive and neurological characteristics of the person being modeled. As the set of parameters of these models have no known connection with each other, and the behavior of such models strongly depends on the values for such parameters, a direct comparison between the models is impossible. Therefore a comparison between the models is made in a more indirect way, by mutual *mirroring* them in each other. This mirroring approach uses any set of values that is assigned to the parameters for one of the models to obtain a number of simulation traces. These simulation traces are approximated by the second model, based on automated parameter estimation. The error for this approximation is considered as a comparison measure. The mirroring is applied in two directions, and also back and forth sequentially by using the estimated parameter values for the second model to estimate new parameter values for the first.

In the paper, first in Section 2 the cognitive model for trust dynamic is described, and in Section 3 simulation results of this model. Section 4 presents a formal analysis of the model. In Section 5 the neural model is presented, with simulation results discussed in Section 6. Section 7 presents a formal analysis of the neural model. In Section 8 the mirroring approach for comparison of models and the automated parameter estimation method are discussed. Finally, Section 9 is a discussion.

2. A cognitive model for relative trust

This section proposes a cognitive model that caters the dynamics of a human's trust on competitive trustees. In this model trust of the human on a trustee depends on the relative experiences with the trustee in comparison to the experiences from all of the competitive trustees. The model defines the total trust of the human as the difference between positive trust and negative trust (distrust) on the trustee. It includes personal human characteristics like trust decay, flexibility, and degree of autonomy (context-independence) of the trust. Fig. 1 shows the dynamic relationships in the proposed model.

In this model it is assumed that the human is bound to request one of the available competitive trustees at each

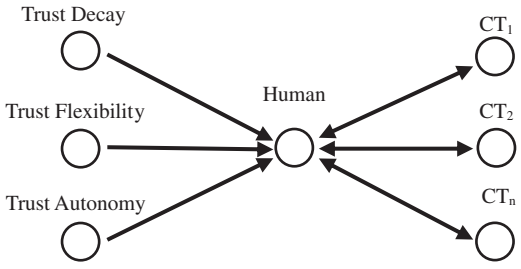

Fig. 1. Trust-based interaction with n competitive trustees.