Trust Model for Human Agent Interaction

Nuno Xu
nuno.xu@tecnico.ulisboa.pt

Supervisors: Rui Prada and Ana Paiva
Técnico Lisboa (Taguspark)
Universidade de Lisboa
Av. Prof. Dr. Aníbal Cavaco Silva
Porto Salvo, Portugal
http://tecnico.ulisboa.pt/en/

1 Introduction

This small document will serve to record the progress on creating a Trust Model for Human Agent Interaction, in the context of my thesis.

2 1st Model

In an effort to create a working trust model iteratively, we will start by simplifying the model described by Castelfranchi and Falcone [1], by removing the effects of outside influence in Trust representation. We also do not take into account long term considerations of the trustor's goal, reducing contextual scope to just the task being performed by the trustor. So Trust is represented by a 3-tuple:

- The trustor (**X**);
- The trustee (**Y**);
- A task ($\boldsymbol{\tau}$) defined by the pair (α, ρ), where $\boldsymbol{\alpha}$ is the action entrusted to the trustee, that possibly produces an outcome $\boldsymbol{\rho}$, contained in the goal of X.

$$TRUST(X\ Y\ \tau) \tag{1}$$

The focus will be in properly representing a certain agent's features that contribute to a trust evaluation, as represented in Figure 1. These features must be able to provide 2 values:

- trust - a value for how much trust we have in this trustee's specific feature;
- certainty - the degree of how much we believe this trust assumption to be true, as factors like how many times, or how long ago, did we last affirm this belief may affect the believability of our assumptions.

Trust and certainty are calculated through a collection of belief sources, which in the first version will be a history of direct contacts. The environment must be able to provide some way to access these contacts, either by a callback function

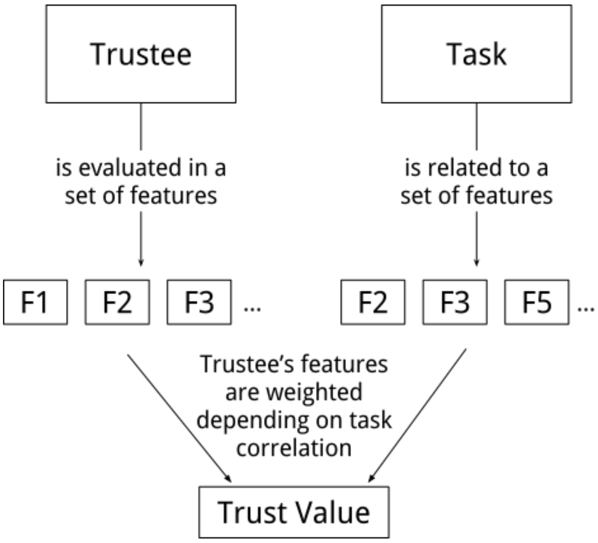
Trustee
Task
is evaluated in a set of features
is related to a set of features
F1
F2
F3
...
F2
F3
F5
...
Trustee's features are weighted depending on task correlation
Trust Value

Figure 1: Trustee Features Representation

or through open access to an action history. If the environment does not provide such features, a perception module, capable of checking action results, will be required.

The task is composed by a set of related features with a given weight. This represents the features most closely related to the task at hand, and their importance to he completion of the task.

2.1 Implementation

Implementation wise, the model will be 1st implemented by using a simple class structure, as seen in figure 2. In this diagram, the main actor is the Agent, which contains a list of Trustees with features that the Agent has been able to perceive from received sources. For now the sources must be given by the simulation environment, but an interpreter should be implemented to sort out and transform the perceptions received by the environment into sources for belief features. A simple simulation example can be performed as following:

1. Instantiate Agent A(nna) and Agent B(ob);
2. Instantiate Trustee B and assign as A's trustee;
3. Insert Direct Contact Source with Feature ID "Cooking"; this should create a new Trust Feature in Trustee B;
4. Calculate Trust with task "Cook";

3 Conclusion

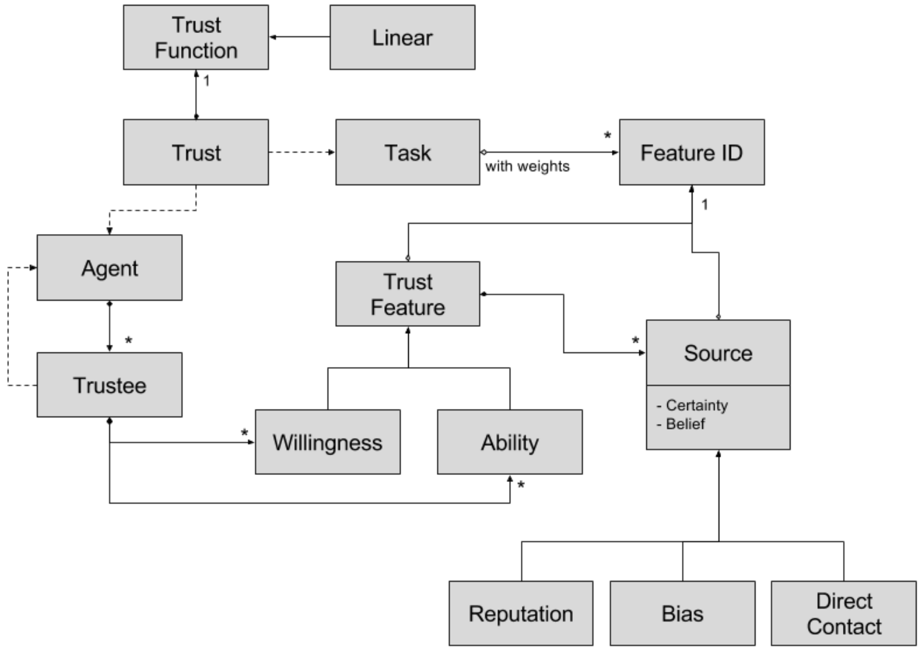

Figure 2: Class Diagram

References

1. Castelfranchi, C., Falcone, R.: Principles of trust for MAS: cognitive anatomy, social importance, and quantification. Proceedings of the International Conference on Multi Agent Systems (1998) 72–79