

Trusting the Messenger Because of the Message: Feedback Dynamics from Information Quality to Source Evaluation

Célia da Costa Pereira

MinD, I3S, Univ. Nice Sophia Antipolis

2015/03/12

ABM

- Simulation using ABM, [Bonabeau, 2002], used for exploring and supporting decision making:
 - these tools (e.g., Repast) provide a graphical development interface and a suite of tools to assist in analysis
 - (actuators) agents are “simple” entities that react to stimulus in the environment
 - suitable for simulations in domains like ecology
 - difficult for simulating human behaviors like in social sciences

BDI

- Belief, Desire, Intention (BDI) model [Rao, Georgeff, 1991]
 - (reasoning) agents are based on a simplified psychological/philosophical view of how people behave
 - pro-active goal seeking behavior
 - BDI programming languages and platforms, JACK [Buseta 1998], JASON [Bordini 2005]

How does the agent adapt its beliefs after receiving a new piece of information?

Accepting a piece of information, [Paglieri & al., 2014]

Belief revision theory:

- in case of inconsistency, give up the old pieces of information
- in absence of inconsistency add the new piece of information
- always keep the new one

The extent to which a reasoner (intelligent) agent considers a message may depend on the:

- message content
- *trust* it has in the source

Trust: a complex concept

Trust as a multidimensional concept

- competence
- source's willingness to convey useful information (e.g., sincerity, malicious, stupid, etc.)

In many real-world situations the trustworthiness is gradual

Trustworthiness and uncertainty

The knowledge of the agent about the capability or the willingness of the source to provide useful information may be incomplete.

How to represent *uncertainty*?

- Working hypothesis: we have a qualitative order on the possible worlds ($\mathcal{I} \in \Omega$)— possibility distribution, π .
- Some worlds (situations) are more possible than others — $\pi(\mathcal{I}_1) \geq \pi(\mathcal{I}_2)$

A Possibility Theory Based Approach

A Possibility Theory Based Approach

Possibility and necessity measures

Definition (Possibility and Necessity Measures)

A possibility distribution π induces a *possibility measure* and its dual *necessity measure*, denoted by Π and N respectively. Both measures apply to a classical set $S \subseteq \Omega$ and are defined as follows:

$$\Pi(S) = \max_{\mathcal{I} \in S} \pi(\mathcal{I}); \quad (1)$$

$$N(S) = 1 - \Pi(\bar{S}) = \min_{\mathcal{I} \in \bar{S}} \{1 - \pi(\mathcal{I})\}. \quad (2)$$

- $\Pi(S)$ expresses to what extent S is consistent with the available knowledge
- $N(S)$ expresses to what extent S is entailed by the available knowledge — impossibility of its complement \bar{S} .

Graded Beliefs

What happens when a new piece of information arrives?

- Each piece of information is understood as a constraint that delimits a set of possible worlds.
- Adding a new piece of information to the agent's belief comes down to discarding worlds that become impossible.
- The more beliefs are available, the smaller the set of possible worlds and the more information is precise.

Definition (Graded Belief)

If π is the normalized possibility distribution π on the possible worlds \mathcal{I} , the degree to which the agent believes ϕ is given by:

$$\mathbf{B}(\phi) = N([\phi]) = 1 - \max_{\mathcal{I} \not\models \phi} \{\pi(\mathcal{I})\}.$$

Modeling competence

We consider a number of competence domains with respect to which the competence of a source s is evaluated:

- the trust in the competence of s is a vector $c(s)$ whose elements may be formally regarded as bipolar degrees of belief $\langle c_d^+, c_d^- \rangle$
- c_d^+ is the degree to which the agent believes the source *is* competent about domain d
- c_d^- is the degree to which the agent believes the source *is not* competent about d

Modeling sincerity

The beliefs an agent maintains about the sincerity of s as a vector of domains $\sigma(s)$:

- bipolar values $\langle \sigma_d^+, \sigma_d^- \rangle$
- $\sigma_d^+ \in [0, 1]$ represents the degree to which the agent has reasons to believe that source s is sincere about d
- $\sigma_d^- \in [0, 1]$ represents the degree to which the agent has reasons to believe the contrary
- $\sigma_d^+ = \sigma_d^-$ represents a status of maximal uncertainty about the sincerity of s about d

Aggregating competence and sincerity

Competence and sincerity may be aggregated into a single trust value τ_d

A source is trusted to the extent that it is believed to be both competent and sincere

For a given domain d , the degree to which s is trusted about d is given by

$$\begin{aligned}\tau_d(s) &= \mathbf{B}(\text{competent}(d, s) \wedge \text{sincere}(d, s)) \\ &= \min(\mathbf{B}(\text{competent}(d, s)), \mathbf{B}(\text{sincere}(d, s))) \\ &= \min(c_d^+(s), \sigma_d^+(s)).\end{aligned}$$

Future work

- other dimensions of trust,
- other categories of source besides sincerity (malicious, intelligent, etc.),
- validate our proposal with real users.