Trustworthy Advice

Nardine Osman and **Patricia Gutierrez** and **Carles Sierra** [1]

Abstract. We propose a novel trust model for assessing the trustworthiness of advice. We say an advice is composed of a plan to execute and a goal to be fulfilled. The expectation of an advice's outcome is calculated by assessing the probabilities of commiting to and executing the plan, and the probability of the executed plan fulfiling the intended goal. The probabilities are learned from similar past experiences.

1 The CONSUASOR Model

'If you go to Ferran Adria's restaurant you will have the time of your life!' 'If you study daily you will get good marks next semester.' These are examples of advice. Advice is what one relies on to help making decisions. But how can one choose which advice to follow and which to discard? CONSUASOR is a trust model that assesses the trustworthiness of advice.

We say an advice has two main components: a plan to execute and the goal to be achieved. In dynamic logic, an advice may be formalised as: $[P_\eta]G$. That is, if η performs plan P, then goal G will be achieved. The question then is: how much should α trust an advice $[P_\eta]G$ recommended by ρ?

CONSUASOR is based on the concept that good advice takes into consideration three issues: (1) the *compliance* of the person being advised with following an advice; (2) the *honour* of the person being advised with respect to executing an advice that he has accepted, and (3) the *causality* describing whether a given plan can cause the intended goal.

We say a trust measure reflects the expectation about an advice's outcome, and we model this expectation as the conditional probability of observing goal G being achieved given ρ's advice $[P_\eta]G$, which we define accordingly:

$$\begin{aligned} &p(Observe(\alpha,G) \mid Commit(\rho,[P_\eta]G)) = \\ &\sum_{P',P'' \in \mathcal{P}} \begin{array}{l} p(Observe(\alpha,G) \mid Observe(\beta,P''_\eta)) \cdot \\ p(Observe(\beta,P''_\beta) \mid Commit(\eta,P'_\eta)) \cdot \\ p(Commit(\eta,P'_\eta) \mid Commit(\rho,[P_\eta]G)) \end{array} \end{aligned} \quad (1)$$

Equation 1 defines this probability as a product of the probabilities of compliance (the probability of η committing to P', given that P was recommended by ρ), honour (the probability of observing η executing P'', given that it committed to P'), and causality (the probability of observing the goal G being achieved, given that the plan P'' was executed). An aggregation is then used for considering all possible plans $\mathcal{P}$ that η may commit to and execute for goal G to be fulfilled.

Of course, different goals may also be achieved. The probability distribution describing all possible outcomes becomes:

$$\begin{aligned} &\mathbb{P}(Observe(\alpha,X) \mid Commit(\rho,[P_\eta]G)) = \\ &\left\{ \begin{array}{l} p(Observe(\alpha,G') \mid Commit(\rho,[P_\eta]G)), \\ p(Observe(\alpha,G'') \mid Commit(\rho,[P_\eta]G)), \\ \ldots \end{array} \right\} \end{aligned} \quad (2)$$

[1] IIIA-CSIC, Spain, email: {nardine,patricia,sierra}@iiia.csic.es

These probabilities are built by learning from past experiences. We define a single experience μ as follows:

$$\langle Commit(\rho,[P_\eta]G), Commit(\eta,P'_\eta), Observe(\beta,P''_\eta), Observe(\alpha,G')\rangle$$

Simplification. We will use the notation $\mathbb{P}_{\mathrm{C}}(X|[P^\rho_\eta]G)$ to describe the probability distribution of compliance $(\mathbb{P}(Commit(\eta,X)|Commit(\rho,[P_\eta]G)))$, $\mathbb{P}_{\mathrm{H}}(X|P_\eta)$ to describe the probability distribution of honour $(\mathbb{P}(Observe(\beta,X_\eta)|Commit(\eta,P)))$, and $\mathbb{P}_{\mathrm{R}}(X|P_\eta)$ to describe the probability distribution of causality $(\mathbb{P}(Observe(\alpha,X)|Observe(\beta,P_\eta)))$. We use the notation $\mathbb{P}_{\mathrm{C|H|R}}(X|\_)$ to refer to any of those three distributions.

Initialisation. At the initial time t_I when no experiences have been considered yet, we choose the uniform distribution $\mathbb{F}$ to describe ignorance: $\mathbb{P}^{t_I}_{\mathrm{C|H|R}}(X|\_) = \mathbb{F}$. As new experiences are encountered, the probabilities are modified as follows.

Update. To update a probability distribution $\mathbb{P}^{t_I}_{\mathrm{C|H|R}}(X|\_)$ with respect to an experience μ, the following is performed:

1. The relevance of the experience μ is calculated accordingly:
   - The relevance of an experience $\mu = \langle Commit(\rho',[P'_\eta]G'), Commit(\eta,P''_\eta), \_, \_\rangle$ with respect to updating $\mathbb{P}^{t_n}_{\mathrm{C}}(X|[P^\rho_\eta]G)$ is calculated as $R_\mu(\mathbb{P}^{t_n}_{\mathrm{C}}(X|[P^\rho_\eta]G)) = (\zeta_g \cdot Sim(G',G) + \zeta_p \cdot Emp(P',P))/(\zeta_g + \zeta_p)$, which describes that an experience is considered relevant if the goals are similar (specified by $Sim(G',G)$) and the past plan empowers the newly recommended plan (specified by $Emp(P',P)$). The similarity between two goals G' and G is measured as their semantic distance in the domain ontology. A plan P' empowers plan P if the capabilities required to execute plan P are implied by the capabilities required for plan P'. The parameters ζ_g and ζ_p help specify the weight of each these measure.
   - The relevance of an experience $\mu = \langle \_, Commit(\eta,P'_\eta), Observe(\beta,P''_\eta), \_\rangle$ with respect to updating $\mathbb{P}^{t_n}_{\mathrm{H}}(X|P_\eta)$ is calculated as $R_\mu(\mathbb{P}^{t_n}_{\mathrm{H}}(X|P_\eta)) = Emp(P',P)$. Note that goals are no longer relevant in this context.
   - The relevance of an experience $\mu = \langle \_, \_, Observe(\beta,P'_{\eta'}), Observe(\alpha,G')\rangle$ with respect to updating $\mathbb{P}^{t_n}_{\mathrm{R}}(X|P_\eta)$ is calculated as $R_\mu(\mathbb{P}^{t_n}_{\mathrm{R}}(X|P_\eta)) = Sim(P',P)$. Note that plan similarity is used as opposed to plan empowerment in this context, as peers' capabilities are not relevant for assessing causal reltions between plans and goals.
2. The experience μ is used to modify the probability of a

single expectation accordingly:

$$p^{t_n}_{\mathrm{C|H|R}}(X{=}x|\_) = \\ p^{t_{n'}}_{\mathrm{C|H|R}}(X{=}x|\_){+}(1{-}p^{t_{n'}}_{\mathrm{C|H|R}}(X{=}x|\_)){\cdot}\epsilon{\cdot}R_\mu(\mathbb{P}_{\mathrm{C|H|R}}(X|\_)) \quad (3)$$

where the past probability $p^{t_{n'}}_{\mathrm{C|H|R}}(X{=}x|\_)$ (calculated at time $t_{n'}{<}t_n$) is increased by a fraction $(\epsilon{\cdot}R_\mu(\mathbb{P}_{\mathrm{C|H|R}}(X|\_)))$ of the total amount that the probability is allowed to increase $(1 - p^{t_{n'}}_{\mathrm{C|H|R}}(X = x|\_))$. This fraction is defined by a fixed percentage that is specified via ϵ, and it is then tuned further by the relevance of μ with respect to x $(R_\mu(\mathbb{P}_{\mathrm{C|H|R}}(X|\_)))$. We note that when assessing compliance or honour, x is chosen based on its semantic distance to what was recommended (P) such that this distance is equal to the semantic distance between P' and P'' of the experience μ (that is, $Sim(P, x) = Sim(P', P'')$). In the case of assessing causality, x is the observed goal G' of μ.

3. The probability distribution is updated w.r.t. the experience μ following the minimum relative entropy approach:

$$\mathbb{P}^{t_n}_{\mathrm{C|H|R}}(X \mid \_) = \underset{\mathbb{P}(X \mid \_)}{\arg\min}\ KL(\mathbb{P}^{t_{n'}}_{\mathrm{C|H|R}}(X \mid \_), \mathbb{P}(X \mid \_)) \\ \text{such that} \quad p(X{=}x \mid \_) = p^{t_n}_{\mathrm{C|H|R}}(X{=}x \mid \_) \quad (4)$$

where KL calculates the Kullbac-Leibler distance, or the relative entropy, between $\mathbb{P}^{t_{n'}}_{\mathrm{C|H|R}}(X \mid \_)$ and the argument $\mathbb{P}(X \mid \_)$, and $p(X{=}x \mid \_) = p^{t_n}_{\mathrm{C|H|R}}(X{=}x \mid \_)$ specifies the constraint that the argument $\mathbb{P}(X \mid \_)$ should satisfy. In other words, we look for distributions that satisfy the newly calculated point $p^{t_n}_{\mathrm{C|H|R}}(X{=}x \mid \_)$ and are at a minimal distance from the original distribution $\mathbb{P}^{t_{n'}}_{\mathrm{C|H|R}}(X \mid \_)$.

Trust Measure. After calculating Equation 2, which we refer to as $\mathbb{P}^{t_n}(X \mid [P^\rho_\eta]G)$ for simplification, the question now is: How do we calculate a *trust measure* based on this expectation that is expressed as a probability distribution? One proposed approach is to calculate the distance between the distribution $\mathbb{P}^{t_n}(X \mid [P^\rho_\eta]G)$ and the distribution representing ρ's promised outcome: $\mathbb{P}_{\mathbf{P}}(X \mid [P_\eta]G) = \{1, \text{if } X = G; 0, \text{otherwise}\}$. The final trust measure is calculated as:

$$trust^{t_n}(\alpha, \rho, [P_\eta]G) = 1 - emd(\mathbb{P}_{\mathbf{P}}(X|[P_\eta]G), \mathbb{P}^{t_n}(X|[P_\eta]G)) \quad (5)$$

where emd measures the earth mover's distance (with the range $[0, 1]$) between two probability distributions.

2 Evaluation

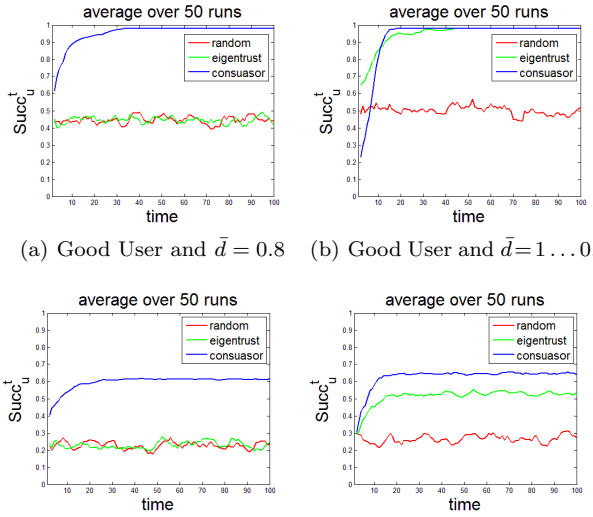
We consider an action meronomy $\mathcal{M}$ and a goal ontology $\mathcal{O}$ of 10 terms each. A set of plans $\mathcal{P}{\in}2^{\mathcal{M}}$ and goals $\mathcal{G}{\in}2^{\mathcal{O}}$. A causality function $f{:}\mathcal{P}{\to}\mathcal{G}$ that describes what goal does each plan achieve. A single user η by the tuple $\langle G_\eta, c_\eta, h_\eta, d_1, d_2\rangle$, where G_η specifies η's goal, $c_\eta{:}\mathcal{P}{\to}\mathcal{P}$ and $h_\eta{:}\mathcal{P}{\to}\mathcal{P}$ describe η's compliance and honour (when P is recommended to η, η commits to $c_\eta(P)$; and when η commits to P, η executes $h_\eta(P)$), tuned by distances $d_1, d_2 \in [0, 1]$ such that $\forall P \in \mathcal{P}$: $Sim(c_\eta(P), P) \geq 1{-}d_1$ and $Sim(h_\eta(P), P) \geq 1{-}d_2$. And a set of recommenders, where each recommender ρ is defined by the tuple $\langle\{c^{-1}_{\rho,\eta}\}_{\forall\eta}, \{h^{-1}_{\rho,\eta}\}_{\forall\eta}, f^{-1}_\rho, d_3, d_4, d_5\}\rangle$ such that:

$$\forall P{\in}\mathcal{P} : Sim(c^{-1}_{\rho,\eta}(P), P') \geq 1 - d_3 \text{ and } c_\eta(P') = P \\ Sim(h^{-1}_{\rho,\eta}(P), P') \geq 1 - d_4 \text{ and } h_\eta(P') = P \\ Sim(f^{-1}_\rho(P), P') \geq 1 - d_5 \text{ and } f(P') = P$$

where $d_3, d_4, d_5 \in [0, 1]$ and $c^{-1}_{\rho,\eta}$, $h^{-1}_{\rho,\eta}$, and f^{-1}_ρ describe what ρ believes the inverse functions of c_η, h_η, and f are.

We compare 3 trust strategies: selecting advise randomly, selecting the advice whose adviser is ranked top by eigentrust [1],[2] and selecting the advice ranked top by CONSUASOR. Each experiment runs for 100 timesteps, and in each timestep: (1) η asks for advice for G_η; (2) *each* recommender ρ suggests a plan $P{=}c^{-1}_{\rho,\eta}(h^{-1}_{\rho,\eta}(f^{-1}_\rho(G)))$; (3) η selects an advice following the experiment's trust strategy, and (4) the experience $\mu{=}\langle P, c_\eta(P), h_\eta(c_\eta(P)), f(h_\eta(c_\eta(P)))\rangle$ is generated, the success of $[P_\eta]G$ is calculated as $Sim(G, f(h_\eta(c_\eta(P))))$, and the current success of η ($Succ^t_\eta$) is an aggregation of the success of all its previously adopted advices.

Figure 1 shows results with good users ($d_1{=}d_2{=}0$) and bad users ($d_1{=}d_2{=}1$). For bad users, $Succ^t_\eta$ remains low because it is not always possible to find plans that achieve G_η and that η is willing to commit to or execute. Cases (a) and (c) have 30 recommenders with medium knowledge ($\bar{d} = 0.8$, where $\bar{d}{=}(d_1{+}d_2{+}d_3)/3$). Cases (b) and (d) have 5 recommenders ranging from fully knowledgeable ($\bar{d}{=}0.6$) to ignorant ($\bar{d}{=}1$). Eigentrust obtains high levels of success when there is at least one good recommender, but when recommenders are not very knowledgeable or users are not fully compliant/honorable, its success diminish. CONSUASOR is able to learn which advices are trustworthy and always reaches high levels of success.



(a) Good User and $\bar{d} = 0.8$ (b) Good User and $\bar{d}{=}1\ldots0.6$

(c) Bad User and $\bar{d} = 0.8$ (d) Bad User and $\bar{d} = 1\ldots0.6$

Figure 1. Success over time

3 Acknowledgments

This work is supported by the Agreement Technologies project (CONSOLIDER CSD2007-0022, INGENIO 2010) and the PRAISE project (EU FP7 grant number 388770).

REFERENCES

[1] Sepandar D. Kamvar, Mario T. Schlosser, and Hector Garcia-Molina, 'The eigentrust algorithm for reputation management in p2p networks', in *Proceedings of WWW-03*, (2003).

[2] In the eigentrust case, we compute the normalized local trust. The notion of transitive trust is not needed since neither users nor recommenders provide advise among themselves.