A Trust Model for a Autonomous Power Management System

July 29, 2014

The content of this paper is an excerpt from [Anders et al. 2013].

Predicting Power Plant Behavior by Means of Trust Values

Because an interaction can last multiple time steps, we define a contract $C_t^i = (c_{t+0}^i, ..., c_{t+n}^i)$ as an (n+1)-tuple that comprises multiple stipulated results c_{t+j}^i, where $j, n, t \in \mathbb{N}_0^+$, $i \in \mathbb{N}^+$ is a unique identifier, and $t + j$ identifies the time step in which the interaction partner should behave as stated in c_{t+j}^i. t and $t + n$ thus specify the time frame in which C_t^i is valid. With respect to power, i.e., residual load, predictions in AVPPs, $n + 1$ is the length of the prediction. For instance, if the power prediction covers a time frame of 8 hours in which the residual load is predicted in 15 minute intervals, we have $n + 1 = 32$. In the following, let $[X]_j$ denote the j-th element of a tuple X. An atomic experience $e_{t+j}^i = (c_{t+j}^i, r_{t+j}^i)$ is a 2-tuple, consisting of the stipulated result c_{t+j}^i and the actual result r_{t+j}^i. An atomic experience $e_{t+j}^i = (7\text{ MW}, 8\text{ MW})$ with an AVPP's residual load states that a residual load of 7 MW was stipulated for time step $t + j$, but 8 MW were measured. Consequently, an experience $E_t^i = (e_{t+0}^i, ..., e_{t+n}^i)$ is an (n+1)-tuple of atomic experiences $e_{t+j}^i = [E_t^i]_j$, and $t+j$ is the time step in which $[E_t^i]_j$ was gained. Contracts C_t^i and experiences E_t^i comprise $n + 1$ so-called *time slots*, e.g., $[E_t^i]_j$ was gained in the j-th time slot.

If an agent a evaluates the trustworthiness of an agent b, it uses a *trust metric* $\mathcal{M} : \mathcal{E} \times ... \times \mathcal{E} \rightarrow \mathcal{T}$ to evaluate a number of experiences with b ($\mathcal{E}$ is the domain of experiences). The metric returns a *trust value* $\tau \in \mathcal{T}$ and relies on a *rating function* $\mathcal{R} : \epsilon \rightarrow \mathcal{T}$ that appraises atomic experiences (ϵ is the domain of atomic experiences). The result of $\mathcal{R}$ is a rating $\pi \in \mathcal{T}$. $\mathcal{T}$ usually is an interval $[0, 1]$ or $[-1, 1]$. Regarding $\mathcal{T} = [0, 1]$, a trust value $\tau = 0$ or $\tau = 1$ states that agent b either never or always behaves beneficially [**?**]. However, b behaves predictably in both cases. If the trust value is around the interval's midpoint, b's behavior is highly unpredictable and thus induces a high level of uncertainty.

Because the residual load can be over- or underestimated, we use $\mathcal{T} = [-1, 1]$ so that positive and negative deviations from predictions can be captured. A rating $\pi = 0$ states that the residual load is predicted exactly, whereas $\pi = -1$ or $\pi = 1$ state that the residual load is greatly under- or overestimated (i.e., the actual residual load is far higher or lower than predicted).

A trust value has to be semantically sound to allow valid predictions of an agent's future behavior. This property depends on the metric $\mathcal{M}$. $\mathcal{M}$ can, e.g., calculate the mean deviation between the stipulated $c^i_{t+j} \in \mathbb{R}$ and actual result $r^i_{t+j} \in \mathbb{R}$ of atomic experiences $[E^{i_h}_{t_h}]_j$ contained in a list of m experiences $E^{i_1}_{t_1}, ..., E^{i_m}_{t_m}$ ($k \in \mathbb{R}$ equals the maximum possible or, if not available, observed deviation from a contract and thus normalizes the result to a value in $[-1, 1]$):

$$\mathcal{M}(E^{i_1}_{t_1}, ..., E^{i_m}_{t_m}) = \frac{\sum_{h=1}^{m} \sum_{j=0}^{n} \mathcal{R}([E^{i_h}_{t_h}]_j)}{m \cdot (n+1)};\ \mathcal{R}([E^i_t]_j) = \frac{c^i_{t+j} - r^i_{t+j}}{k} \tag{1}$$

Based on $\mathcal{M}$, a trust value τ, and a contract C^i_t, an agent can predict the *expected behaviors* $B^i_t = (b^i_{t+0}, ..., b^i_{t+n})$ of its interaction partner during C^i_t's validity. With respect to Eq. 1, the agent's *expected behavior* $[B^i_t]_j$ in time step $t + j$ is defined as the difference between $[C^i_t]_j$ and the expected deviation $\tau \cdot k$:

$$[B^i_t]_j = [C^i_t]_j - \tau \cdot k \tag{2}$$

For example, if $k = 10$ MW, $\tau = 0.1$, and the power prediction's stipulated results are $C^i_t = (5 \text{ MW}, 6 \text{ MW})$, the expected residual load can be predicted as $B^i_t = (4 \text{ MW}, 5 \text{ MW})$. If the AVPP schedules its subordinate controllable power plants on the basis of B^i_t instead of C^i_t, it is expected that the deviation between the power plants' output and the actual residual load can be decreased. However, the prediction of the residual load's future behavior with Eq. 2 can be imprecise because we disregard that an agent's behavior can be arbitrary and that it might be time-dependent. Since agents can behave arbitrarily, one and the same trust value can stem from very different experiences, e.g., $\tau = 0.1$ could be based on experiences in which the residual load was always 1 MW lower than stipulated or a situation in which 25% of the predictions were overestimated by 2 MW and 75% of the predictions were underestimated by -2 MW. With regard to time-dependent behavior, the prediction of the residual load for a time step t could, e.g., tend to be rather precise if the prediction for the previous time step $t - 1$ is accurate. A similar dependence could exist for inaccurate predictions.

References

[Anders et al. 2013] Gerrit Anders, Florian Siefert, Jan-Philipp Steghöfer, and Wolfgang Reif. 2013a. Trust-Based Scenarios – Predicting Future Agent Behavior in Open Self-Organizing Systems. In *Self-Organizing Systems – Proc. of the 7th International Workshop on Self-Organizing Systems (IWSOS 2013) (Lecture Notes in Computer Science)*, Wilfried Elmenreich, Falko Dressler, and Vittorio Loreto (Eds.), Vol. 8221. 90–102.