Dynamically Maintaining Standards using Incentives

Ramón Hermoso[1] and Henrique Lopes Cardoso[2]

Abstract Standards have had much importance in different fields of research in order to assure a certain quality of service in bilateral contracts. More specifically, in multi-agent systems performance standards may be used in order to articulate contracts among partners in environments dealing with uncertainty. However, little effort has been made on how to ensure standards compliance over time. In this work we put forward a learning-based mechanism that attempts to maintain performance standards by applying incentives and/or punishments to agents identified as specialised for certain tasks. We present some empirical results supporting our approach.

Key words: multiagent systems, standard, incentive

1 Introduction

A number of research proposals have been made recently concerning the development of infrastructures for supporting interaction in open multi-agent systems. In such systems agents enter and leave the interaction environment, and behave in an autonomous and not necessarily cooperative manner, exhibiting self-interested behaviours. Even when agents establish commitments among them, the dynamic nature of the environment may jeopardize such commitments if agents are not socially concerned enough, valuing more their private goals when evaluating the new circumstances.

Moreover, in open systems one cannot assume that agents will behave consistently along time. This may happen either because of agent's ability or

[1]CETINIA, University Rey Juan Carlos - Tulipán s/n, 28933, Madrid, Spain
e-mail: ramon.hermoso@urjc.es
·[2]LIACC / DEI, Faculdade de Engenharia, Universidade do Porto - Rua Dr. Roberto Frias, 4200-465 Porto, Portugal e-mail: hlc@fe.up.pt

benevolence. In some cases, an agent may not be capable of maintaining a certain behaviour standard throughout its lifetime. In other cases, the agent may intentionally deviate from its previous performance. It is therefore important, when considering open environments, to take into account also the evolution of an agent's internal skills or motivations, besides the dynamics of the interaction environment as a whole.

Looking at the society from a role-specialization perspective, Hermoso *et al.* [3] propose a coordination mechanism, based on role evolution, that assists agents in selecting good partners to whom to delegate specific tasks. The authors look at the agent society and identify "run-time roles" that cluster agents with similar skills for (sets of) tasks. This allows one to identify the role that labels agents most suitable to perform a specific task.

Building on this work, in this paper we associate roles with *performance standards* and address their maintenance: given the dynamics of agents' behaviour, how can performance standards be guaranteed? Two different policies can be used when agents start under-performing. One is to update the role taxonomy and to measure new standards. But assuming that this reorganization may be costly, another option is to influence agents' reasoning by employing incentives, as an attempt to keep them on track.

In Section 2 we put forward a model to establish and adjust incentives in order to maintain standards over time. We present some empirical results in Section 3. Finally, we sum up the paper and point future work in Section 4.

2 Incentive-based mechanism to maintain standards

While the work in [3] focused on providing a role specialization taxonomy enabling better trust estimations of agents when performing specific tasks, in this paper we assume that such roles may be used to assess performance standards that provide a clearer picture of agents' skills. The path from roles to standards is described in [4]. These measured standards are then to be maintained through an incentive-based policy.

In order to devise an incentive mechanism, we model our interaction scenario according to the well known *principal-agent* model [5, 1] from economics, in which a principal (a service requester) requests an agent (the provider) to perform a specific task. The principal is interested in influencing the efforts that the agent puts when performing the task: efforts correspond to available actions with different execution costs. The exact actions executed by the agent are unobservable to the principal; instead, the latter observes some performance measures. Actions determine stochastically the obtained performance, which is therefore a random variable whose probability distribution depends on the actions taken by the agent. By establishing an incentive schedule, the principal aims at encouraging the agent to choose actions better leading to an intended performance standard.

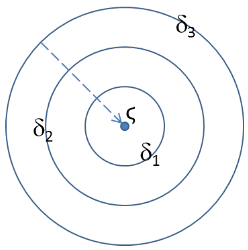


Fig. 1 A standard as a target

2.1 Targeting standards

Standards are generated using an averaging function applied to task execution outcomes of a group of provider agents (as explained in [4]). Since, according to our model, standards allow requesters to identify expected values for task executions, we consider a standard as a target that agents should meet. Any deviation from the standard is seen as a sub-optimal outcome. Figure 1 illustrates this idea, where ς represents the expected target standard, and each concentric circle labelled with δ_i denotes equidistant performances to the target. These concentric lines highlight the fact that deviations in any direction are considered equally harmful in terms of expected values. The arrow pointing towards the centre discloses the aim of our incentive-based approach: to encourage providers to better target the standard.

We assume each provider has a set of actions at its disposal, each with a cost and a probability function for obtaining different performance outcomes. As follows from Figure 1, an outcome is seen as a *distance* to the standard. This allows us to think of actions as *efforts* the provider puts in when executing a given task: the more effort is invested, the higher the likelihood that the obtained outcome will be closer to the standard. Naturally, expending more effort also means bearing a higher cost.

2.2 Actions, outcomes and incentives

More formally, using a finite model for actions and outcomes, we have that:

- The provider has an ordered set of possible actions $\mathcal{A} = \{a_1, ..., a_n\}$, where $a_i \prec a_j$ if $i < j$. This means that $Cost(a_i) < Cost(a_j)$.
- The possible observable outcomes the provider may obtain is an ordered set $\bar{\mathcal{X}} = \{\bar{x}_1, ..., \bar{x}_m\}$, where $\bar{x}_i \prec \bar{x}_j$ if $i < j$ ($\bar{x}_i$ is a worse performance than $\bar{x}_j$). For simplification, we assume that $\bar{x}_i \in [0, 1]$, for all $i \in [1, m]$: each $\bar{x}_i$ denotes the percentage of the target standard that has been achieved.

- There is a probability distribution function for $\bar{\mathcal{X}}$ given an action in $\mathcal{A}$, where $p(\bar{x}_k|a_i)$ is the probability of obtaining outcome $\bar{x}_k \in \bar{\mathcal{X}}$ when performing action $a_i \in \mathcal{A}$. We have that $\sum_{k=1}^{m} p(\bar{x}_k|a_i) = 1$, for all $i \in [1, n]$.

We assume that the monotone likelihood ratio property (MLRP) [1], relating actions with outcomes, holds for every provider. MLRP states that greater efforts are more likely to produce better outcomes: for any $a_i, a_j \in \mathcal{A}$ with $a_i \prec a_j$, the likelihood ratio $p(\bar{x}_k|a_i)/p(\bar{x}_k|a_j)$ is non-increasing in k.

Incentives are specified through an incentive schedule function mapping possible outcomes to values to be collected or paid by the provider: $I : \bar{\mathcal{X}} \to \mathcal{I}$. We take I to be non-decreasing, that is, $I(\bar{x}_1) \leq ... \leq I(\bar{x}_m)$, meaning that higher outcomes must have at least the same incentive as lower ones. Moreover, we look at incentives as producing some change in the utility the agent would get if no incentives were in place; in this sense, $\mathcal{I} = \{\iota : \iota \in [-1, 1]\}$, where positive (negative) values denote percentage increases (decreases) in utility. When $\iota = 0$ there is no incentive in place.

Based on the stochastic model of action outcomes explained above, each provider is taken to be expected utility maximizer. Therefore, when choosing the action to perform it will maximize expected utility [9]:

$$\arg\max_{a \in \mathcal{A}} \mathbb{E}_a = \sum_{i=1}^{m} p(\bar{x}_i|a) u(I(\bar{x}_i)) - Cost(a) \tag{1}$$

where $u(I(\bar{x}_i))$ is the utility the agent gets from obtaining performance outcome $\bar{x}_i$ and consequently incentive $I(\bar{x}_i)$. Function $u : \mathcal{I} \to [0, 1]$ is taken to be strictly increasing. We assume provider agents are risk averse. We define function u using a sigmoid:

$$u(I(\bar{x})) = \frac{1}{1 + e^{-I(\bar{x}) \cdot B + \kappa}} \tag{2}$$

where $\kappa \in \mathbb{R}$ represents a parameter to tune the center of the sigmoid function and $B \in \mathbb{N}^+$ allows us to tune the sensitivity to received incentives.

2.3 Deviations and responses

Given the previous performance of each provider, on which standards have been defined, we identify two possible causes for agents to deviate from those standards, in the sense that they are not able to meet them anymore. Such causes naturally come to surface from analysing Equation 1: i) action costs have changed, leading an agent to choose actions that stochastically obtain lower outcomes; ii) probabilities for an action's performance outcomes have changed, e.g. due to environmental factors not under the control of the agent, meaning that a specific action is not as effective as before.

These deviations in performance may make a role taxonomy and its previously measured performance standards inaccurate to represent agents' current capabilities. In order to maintain standards, the system may determine and employ an appropriate incentive schedule $I : \bar{\mathcal{X}} \rightarrow \mathcal{I}$, which is based on measurable outcomes of task execution. As mentioned before, an outcome is a percentage of the target standard that has been met. Unlike typical approaches in game theory, we do *not* assume any knowledge of the incentive policy maker regarding action costs and probability distributions over outcomes, or provider utility functions. Thus, we see the problem of searching for an optimal incentive schedule as a *reinforcement learning* (RL) [8] problem.

In the following we briefly describe how states, actions and rewards are addressed in the problem faced by the incentive policy maker.

States. The state entails recently obtained performance outcomes. States exhibiting performances farther away from the target standard need to be addressed with stronger incentive policies, while states denoting abidance to agreed standards need no intervention from the policy maker.

Depending on how performance quality is to be interpreted, we may aggregate recent task executions in different ways. In this paper we rely on an average: $perf = \left(\sum_{i=t-\Delta}^{t} \bar{x}^i\right)/\Delta$, where t is the current time step, $\bar{x}^i$ is the outcome obtained at time step i and Δ is the size of the time window, i.e. the number of task executions to consider.

In order to reduce the size of the state space, states are discretized according to the number of levels of deviation that are to be addressed differently, as illustrated in Figure 1. We define a δ parameter specifying in how many intervals to split the distance to target standards:

$$state = \begin{cases} 1 & \text{if } perf = 1 \\ \lfloor perf \cdot \delta \rfloor / (\delta - 1) & \text{if } perf < 1 \end{cases}$$

This function gives us δ different states, represented by values within $[0, 1]$.

Actions. Available learner actions concern incentive schedules I that specify, for any $\bar{x} \in \bar{\mathcal{X}}$, an incentive value $\iota \in \mathcal{I}$. Following [1], each action can be seen as a non-decreasing incentive vector $(\iota_1, ..., \iota_m)$, where m is the number of possible outcomes. In order to reduce the action space, we consider only incentive values in the set $\lfloor \mathcal{I} \cdot 10 \rfloor / 10$ (discrete values with 0.1 steps). Yet, depending on the number of outcomes to consider, this may still give us a huge number of actions to experiment with.

The heuristic we use to tackle with this problem is to explore the action space by generating incentive schedules that consist of minor changes to the currently employed schedule: we step-change one of the incentive values and if needed fix the rest of the schedule to guarantee the non-decreasing property. A *softmax* policy [8] is used to select among the actions considered.

Rewards. An optimal incentive schedule should take into account both the obtained provider outcomes and the cost of applying the incentive schedule. In our approach, these costs are associated with the actual performances that such a schedule has led to, since incentives are paid (if positive) or collected (if negative) according to actual outcomes. Considering that the mechanism does not seek profit, but rather to intervene as least as possible, we sum the absolute values of actually applied incentives when computing the incentive schedule cost.

A reward is computed as a weighted difference between the sum of obtained outcomes and the cost of the incentive schedule. Using weights allows us to define the relative importance of providers' performance and incentive cost.

In RL, $Q(s, a)$ values are computed to determine the expected return for executing action a in state s. We update these values using the simple update rule $Q(s, a) = Q(s, a) + \alpha \cdot (reward - Q(s, a))$, where α is a step-size parameter (we use $\alpha = 0.3$ for the following experimental evaluation).

3 Experiments

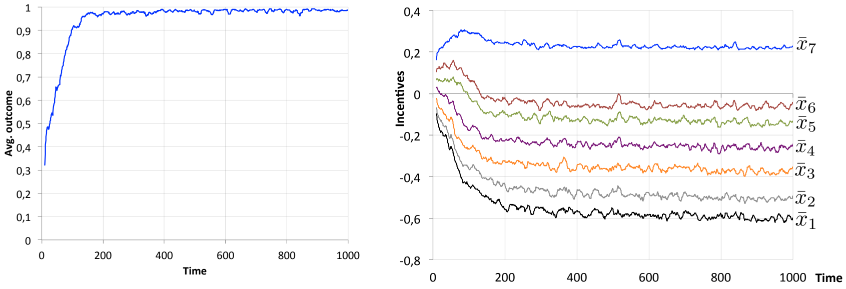
We have implemented a simulation environment by using the Repast framework. In order to calculate actual outcomes when a task is requested, providers' behaviour is defined in terms of possible outcomes. In order to do that, we need to set a relationship between efforts and actual outcomes. We have modelled this issue by using beta distributions. There exists a different beta distribution for every different possible effort, in order to be able to calculate actual outcomes. We consider as possible outcomes the set $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_7$, where $\bar{x}_i$ are different equidistant values in $[0, 1]$. For the sake of simplicity, the number of different efforts available to providers is the same (although it needs not be): $a_1, a_2, \ldots, a_7$. We set the centre value for each effort as the outcome value with the same index: each effort a_i will obtain an outcome modelled as a beta distribution centred in $\bar{x}_i$. In the experiments reported in this paper, all providers share the same beta distributions.

We also need to define a function for effort costs. These costs are used in the provider's decision making (see Equation 1). For that purpose, we use Equation 3 to define different profiles of providers. This means that different providers may have different costs for the same efforts.

$$Cost(a) = \alpha \cdot (\rho + (1 - \rho) \cdot a^{1/\beta}) \tag{3}$$

In this set of experiments we have a heterogeneous population of 100 providers, with random values for α, β and ρ, thus obtaining individuals with a different curve relating efforts to their costs. We set values κ and B to 0 and 10, respectively (see Equation 2). Those values fix the sensibility of providers to incentives in their decision-making processes.

We simulate task requests from customers, one to every different provider in every time step. We show average results from 10 different runs.



(a) Average outcomes

(b) Incentive schedule evolution

Fig. 2 Experimental results

In Figure 2(a) we observe how our approach progressively learns an appropriate incentive schedule, which induces providers to behave better: they get progressively closer to the standard. We can also see that this process takes some time, since there exist a high number of possible new (unexplored) incentive schedules that can be generated in each step of the learning process. Once the approach converges to an (almost) optimal achievement of the standard, an appropriate incentive schedule makes providers select the most reliable action (in terms of standard achievement).

Figure 2(b) shows the evolution of the incentive schedule employed. Incentives applied to each possible outcome are shown.

4 Conclusions and future work

Standards are used as a means to articulate contracts in social interactions. In this paper we have proposed a mechanism that provides incentives to make agents maintain a level of performance as close as possible to the standards. Some possible applications of this approach cover from manufacturing systems, in which agents playing different roles when building a craft are supposed to meet and maintain a standard during their work, to social systems such as ruled electronic markets, where while standards may not be known a priori, they can be discovered at runtime and artificially maintained for the sake of the overall market community.

There are economic approaches also founded on the emergence of standards. Sherstyuk [7] proposes a method to set appropriate performance standards to develop optimal contracts, in which the provider's best choice is to keep the standard through its action. In this paper, however, we are not pursuing optimal performance standards; instead, we are concerned about how to maintain the level of those standards once they have been created.

In the same line Centeno *et al.* [2] present an approach on adaptive sanction learning by exploring and identifying individuals' inherent preferences without explicit disclose of information – the mechanism learns over which attributes of the system should modifications be applied in order to induce agents to avoid undesired actions. In our case, we adhere to a more formal scenario, in which interactions are regulated by means of contracts. Moreover, we assume that the attributes that may be modified by means of incentives are already known by the mechanism.

The approach taken in [6] also assumes that the mechanism knows which attributes it should tweak in order to influence agents' behaviors, namely by adjusting deterrence sanctions applicable to contractual obligations that agents have committed to. The notion of social control employed there is similar to our notion of role standard maintenance; however, instead of a run-time discovered standard, a fixed threshold is used to guide the decisions of the policy maker. Moreover, only sanctions (seen as fines) are used to discourage agents from misbehaving, while here we are also interested in incentivating agents to do their best (by using appropriate actions) while executing the tasks they are assigned to.

We intend to pursue the mechanism presented in this paper, namely by refining the learning model of the incentive policy maker. We also intend to combine the approach with the decision on when to reconfigure the role taxonomy from which standards have been generated.

References

1. B. Caillaud and B. Hermalin. Hidden action and incentives. Teaching Notes, U.C. Berkeley, accessed at http://faculty.haas.berkeley.edu/hermalin/agencyread.pdf, 2000.
2. R. Centeno, H. Billhardt, and R. Hermoso. An adaptive sanctioning mechanism for open multi-agent systems regulated by norms. In *Proc. of the 23rd IEEE Int. Conf. on Tools with Artificial Intelligence*, pages 523–530. IEEE Computer Society, 2011.
3. R. Hermoso, H. Billhardt, and S. Ossowski. Role evolution in open multi-agent systems as an information source for trust. In *9th International Conference on Autonomous Agents and Multi-Agent Systems*, pages 217–224. IFAAMAS, 2010.
4. R. Hermoso and H. Lopes Cardoso. Dynamic discovery and maintenance of role-based performance standards. In Ossowski, Toni, and Vouros, editors, *Agreement Technologies*, volume 918 of *CEUR Workshop Proceedings*, pages 27–41. CEUR-WS.org, 2012.
5. J.J. Laffont and D. Martimort. *The Theory of Incentives: The Principal-Agent Model.* Princeton paperbacks. Princeton University Press, 2002.
6. H. Lopes Cardoso and E. Oliveira. Social control in a normative framework: An adaptive deterrence approach. *Web Intelligence and Agent Systems*, 9:363–375, December 2011.
7. K. Sherstyuk. Performance standards and incentive pay in agency contracts. *Scandinavian Journal of Economics*, 102(4):725–736, 2000.
8. R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction.* The MIT Press, 1998.
9. J. Von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior.* Princeton University Press, 3 edition, May 1980.