Artificial Intelligence 186 (2012) 38–94

Contents lists available at SciVerse ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Communicating open systems

Mark d'Inverno [a], Michael Luck [b], Pablo Noriega [c], Juan A. Rodriguez-Aguilar [c,*], Carles Sierra [c]

[a] Department of Computing, Goldsmiths, University of London, London SE14 6NW, United Kingdom
[b] Department of Informatics, King's College London, Strand, London WC2R 2LS, United Kingdom
[c] Artificial Intelligence Research Institute, IIIA-CSIC, Spanish National Research Council, Campus de la Universitat Autònoma de Barcelona, 08193 Bellaterra, Catalonia, Spain

ARTICLE INFO

Article history:
Received 5 July 2011
Received in revised form 8 March 2012
Accepted 11 March 2012
Available online 13 March 2012

We dedicate this work to the memory of our friend and colleague, Marc Esteva, who sadly passed away in December 2011[1]

Keywords:
Open systems
Agent communication
Formal languages
Electronic institutions

ABSTRACT

Just as conventional institutions are organisational structures for coordinating the activities of multiple interacting individuals, electronic institutions provide a computational analogue for coordinating the activities of multiple interacting software agents. In this paper, we argue that open multi-agent systems can be effectively designed and implemented as electronic institutions, for which we provide a comprehensive computational model. More specifically, the paper provides an operational semantics for electronic institutions, specifying the essential data structures, the state representation and the key operations necessary to implement them. We specify the agent workflow structure that is the core component of such electronic institutions and particular instantiations of knowledge representation languages that support the institutional model. In so doing, we provide the first formal account of the electronic institution concept in a rigorous and unambiguous way.

1. Introduction

Open systems [61], in which the various constituent components are unknown in advance and can change over time, are increasingly becoming a *de facto* model for computing. Not only do they reflect the need for interconnection and interaction that are required by modern information systems, they also underpin several visions of future computing systems that span grid computing [52], ambient intelligence [92] and the semantic web [12], as well as many others. Such systems are characterised by: decentralised control, avoiding the bottleneck of a centralised decision-maker; concurrency, by which the different components operate simultaneously with others; and loose coupling, with no component having access to the internal state or structure of others.

At the same time, multi-agent systems have emerged as a promising approach for the development of agile information systems, and are well suited to addressing problems that have multiple problem-solving methods, multiple perspectives or multiple problem-solving entities [66]. In addition to inheriting the traditional advantages of distributed problem-solving, multi-agent systems are based on the exploitation of numerous varieties of sophisticated patterns of interaction, enabling agents to engage in many distinct forms of behaviour. For example, agents may cooperate to achieve a common goal, they

* Corresponding author.
E-mail addresses: dinverno@gold.ac.uk (M. d'Inverno), michael.luck@kcl.ac.uk (M. Luck), pablo@iiia.csic.es (P. Noriega), jar@iiia.csic.es (J.A. Rodriguez-Aguilar), sierra@iiia.csic.es (C. Sierra).

[1] Many of the ideas found in this paper stem from a close and intense collaboration with Marc over the last decade. We have valued his insight and his intellectual contributions to our joint work, to the research community, and to the research domain of multi-agent systems. Most of all, however, we have valued his friendship. The wealth of joyful moments we have shared with Marc over the years will remain constant in our memories. We miss him and will continue to do so.

0004-3702/$ – see front matter
http://dx.doi.org/10.1016/j.artint.2012.03.004

may coordinate their activities in order either to avoid harmful interactions or to exploit beneficial interactions, or they may negotiate agreements. Such varieties of interaction provide the means for multi-agent systems to be highly flexible, in a very different fashion to other forms of software. However, the design and development of such multi-agent systems suffer from all the problems associated with the development of distributed concurrent systems and the additional problems that arise from the kinds of flexible and complex interactions envisaged among autonomous entities [66]. Moreover, the complexity of designing multi-agent systems increases when they are also open systems.

Such open multi-agent systems are populated by heterogeneous agents, and can be considered to be developed by different people using different languages and architectures, representing different parties, and acting to achieve individual goals. Since they are highly complex, costly and may sustain critical applications, it is vital to adopt principled methodologies that support their specification, analysis and validation [66,11]. Indeed, there has been a surge of interest in agent-oriented methodologies and modelling techniques in recent years, motivated and driven both by work on the development of first generation agent systems, which have informed subsequent efforts, and by the need to address the concerns raised in seeking to deliver the visions of future computing systems, as suggested above.

While much work has focused on the micro-level (agent-centred) view, in which the control architecture of individual agents is the key concern, the macro-level (organisation-centred) view of multi-agent systems requires equal attention, particularly in light of the demands for interconnection and interaction. Indeed, there has recently been increasing interest in incorporating organisational concepts into multi-agent systems as well as in shifting from agent-centred to organisation-centred designs [10,34,40,46,78,85,95,100] that treat the organisation as a first-class citizen, similar to the views articulated in pioneering work by Gasser [58] and Pattison et al. [87].

Here, organisations structure the activities of the entities involved, or control the actions of a system as a corporate entity. In this view, a shared organisational structure provides agents with descriptions of their roles and responsibilities in the multi-agent context and contains guidelines for their intelligent cooperation and communication. In other words, an organisational structure defines a behaviour space for agents with a set of conventions, or rules of behaviour, that agents are required to follow. Of course, many different types of organisational structure (that specify the roles played by the various agents in the system, their activities, relationships among these activities, and so on) are possible, and providing a means by which these can be specified and constructed is important. One way of providing such organisational structure for open systems is through electronic institutions that provide a computational analogue of conventional institutions.

According to North [80], human interactions themselves are guided by institutions, which represent the rules of the game in a society, including any (formal or informal) kind of constraint that humans devise to shape their interactions. Thus, institutions are the framework within which human interactions take place, specifying what individuals are forbidden from doing and permitted to do, the sanctions that may be imposed if they do not comply, and under what conditions. Human organisations and individuals conform to these institutional rules in order to receive legitimacy and support.[2] In this view, establishing a stable organisational structure for human interactions provides the *raison d'être* of institutions.

The successful adoption of institutions by human societies as a means of structuring and regulating interactions suggests that we might also use institutional notions for structuring and regulating computational interactions, in particular to cope with the complexity of deploying open multi-agent systems. Importantly, this inspiration from human institutions, for use in institutions of humans and agents or even entirely computational agent institutions, in the management and regulation of traditional information systems, for example, demands that we incorporate several key principles. First, the context of any institution must be explicitly taken into consideration as a means of constraining interactions; interactions between individuals are constrained by the prior history that has led them to their current position, and actions can only be situated in that context. Moreover, institutional interactions are persistent in the sense that they are not one-shot computations, but provide the overall institutional context in which other actions can be taken. Actions have such social persistence so that even if an individual agent does not remember what has happened previously, there are consequences for the institution as a whole, and for the individuals within it. In addition, the interactions that underpin the operation of an institution are not necessarily just between two agents, as has often been the restricted case considered in the literature, but instead are collective, involving multiple agents in different roles. Finally, subject to these constraints, each agent is autonomous, and makes individual decisions about actions in light of the context and persistence of the institution. Such decision making impacts on the decisions and actions of others subsequently, and such decision making must therefore be appropriately signalled to others.

In this paper, we argue that open multi-agent systems can be effectively designed and implemented as electronic institutions, for which we provide a comprehensive computational model. The paper makes several distinct contributions, as follows.

First, the paper provides a *reference ontology* for agents engaging in interaction as envisaged here. The paper thus establishes a framework of terminology to describe a range of open systems.

Second, we the paper gives data structures and operations that constitute core *governance mechanisms* to regulate autonomy of agents.

Third, the paper provides a *formal operational specification* for electronic institutions in support of two distinct aims: first, to provide a clear and unambiguous exposition of the concepts underlying electronic institutions; and second, to enable

[2] For instance, *Robert's Rules of Order* (of parliamentary procedure) is a well-known example of an institution.