Creating and Reusing Metric Graphs for Evaluating Agent Performance in the Supply Chain Management domain

Christos Dimou[1], Kyriakos C. Chatzidimitriou[1], Andreas L. Symeonidis[1,2], and Pericles A. Mitkas[1,2]

[1] Electrical and Computer Engineering Dept., Aristotle University of Thessaloniki, GR541 24, Thessaloniki, Greece
[2] Intelligent Systems and Software Engineering Laboratory, Informatics and Telematics Institute – CERTH, GR570 01, Thessaloniki, Greece
cdimou@issel.ee.auth.gr, kyrcha@issel.ee.auth.gr, asymeon@iti.gr, mitkas@eng.auth.gr

Abstract. The overwhelming demand for efficient agent performance in Supply Chain Management systems, as exemplified by numerous international competitions, raises the issue of defining and using generalized methods for performance evaluation. Up until now, most researchers test their findings in an ad-hoc manner, often having to re-invent existing evaluation-specific knowledge. In this position paper, we tackle the key issue of defining and using metrics within the context of evaluating agent performance in the SCM domain. We propose the Metrics Representation Graph, a structure that organizes performance metrics in hierarchical manner, and perform a preliminary assessment by instantiating an MRG for the TAC SCM competition, one of the most demanding SCM competitions currently established. We envision the automated generation of the MRG, as well as appropriate contribution from the TAC community towards the finalization of the MRG, so that it will be readily available for future performance evaluations.

1 Introduction

During the last decade, we have witnessed an impressive shift in commercial transactions from traditional practices to alternatives that are mainly based on Information and Communication Technologies (ICT) infrastructures. Among the different technologies employed, agents and multi-agent systems prove to be suitable for tasks such as employing trading strategies, participating in electronic auctions and managing the assets and the resources of their holders. Indeed, the nature of e-commerce fully reveals and exploits the primary characteristics of agent technology, such as autonomy, proactiveness and adaptivity [1].

From a software engineering perspective, it is evident that the need for evaluating the performance and effectiveness of the proposed methods is imperative. However, in practice, most researchers conduct their evaluation sessions in an ad-hoc manner, being obliged to define from scratch appropriate metrics and

measurement methods. Their findings are not reproducible, while the extra effort of re-inventing existing knowledge is burdensome, both in terms of time and cost.

Driven by the need of organizing and reusing evaluation-specific knowledge, we propose the Metrics Representation Graph (MRG), a structure that organizes metrics in an hierarchical manner and makes them readily available for evaluators within a specific domain (Section 3). In this paper, we raise and discuss issues with respect to both implementing and using the proposed MRG to the Trading Agent Competition [3] on Supply Chain Management (TAC SCM) [2], a popular, annual event that encourages research towards more effective trading agents within the context of complex SCM scenarios (Section 4). We continue with presenting the related background (Section 2) and conclude with a section on future work (Section 5).

2 Background

The problem of evaluating and assessing the performance of intelligent systems (agent-based systems being a specific category of them), is a crucial, yet overlooked aspect of modern computing. Indeed, there is a remarkable lack of methodologies and tools that address the problem of evaluation in a generalized and coherent fashion. Researchers often develop their own *ad-hoc* evaluation methods and tools for every newly proposed agent-based system they develop, making practically impossible for third parties to repeat and validate their findings. Moreover, each evaluator is required to determine various evaluation aspects (such as appropriate domain metrics) from scratch. For example, different evaluation methodologies have been used in TAC SCM with the most eminent that of ad-hoc controlled experiments that test versions of an agent with itself and other strawmen agents [3, 4]. In [5] an effort is made towards controlling the complexity of experiments and for agent testing and analysis. In order to assist evaluators through a set of standardized steps and representation tools, the Agent Performance Evaluation (APE) methodology has been proposed in [6]. The core methodology comprises three main sections: a) metrics representation, b) measurement collection and c) aggregation. In the remainder of this paper, we focus on metrics representation and we demonstrate the organization and reuse of evaluation-specific information in the TAC SCM domain.

3 Knowledge reuse for evaluation purposes

Within the context of APE, we propose the Metrics Representation Graph (MRG), a representation tool for defining and associating metrics for a specific application domain. The main objectives of the MRG are to:

1. provide a complete set of measurable and abstract metrics for a specific application domain

[3] http://www.sics.se/tac

2. define all metric-related information, including scales of measurement, cardinality, etc., for each metric
3. organize and associate the above metrics in a weighted, hierarchical manner

The basic layering of an MRG is depicted in Fig. 1. Each node of this acyclic graph represents a metric. The bottom-most level of the graph comprises *Simple Metrics*, i.e. metrics that are directly measurable from experiments. Moving upwards in the MRG, Simple Metrics are combined into *Composite Metrics*, i.e. metrics that represent higher level concepts that are not directly measurable (such as robustness and durability). An MRG may have any number of levels that combine Composite and/or Simple Metrics, upwards to the root node, the overall *System Performance* metric. The edges of the graph represent the weighted contribution of each child metric to its parent metric, a required information for the subsequent *aggregation* of collected measurements.

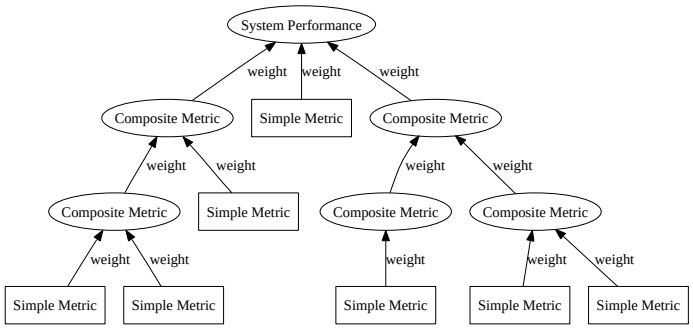

Fig. 1. The basic structure of the Metrics Representation Graph

It is evident that the MRG is a flexible structure that maintains and makes readily available all metrics-related information for a specific application domain. In other words, a beginner evaluator in a specific domain is able to use the corresponding MRG *off-the-shelf* in the context of his/her application, while being able to select branches of the MRG, in order to evaluate the application at hand at different levels of granularity, from coarser to finer grain of metrics detail.

4 A Metrics Representation Graph for the TAC SCM domain

In order to demonstrate the added-value of the APE methodology we have applied it on the SCM domain and particularly on the TAC SCM simulation framework. Within the TAC SCM game each agent acts as a Personal Computer (PC) manufacturer, competing against other five agents on supplier and customer

contracts. Throughout the duration of the game, each agent has to: (a) negotiate supply contracts, (b) bid for customer orders, (c) manage daily assembly activities and, (d) ship completed orders to customers. The performance measure used to rank agents throughout the competition is their final bank account at the end of 220 simulated days of activity.

The performance of an agent may be affected by a set of environmental factors (for example supply and demand fluctuations) simulated by the TAC SCM server, by its competitors and the agent's strategy. Based on the facts that we want to create a MRG to help us during game-playing while we cannot have access to information regarding other agents at runtime (usually the case in real-life scenarios), we have focused our work on measurable metrics from the viewpoint of the agent under investigation. The profile space [7], i.e. the influence of the mix of competing agents, is introduced implicitly through some metrics provided by reports during the game. Taking all the considerations into account, we have developed the MRG for TAC SCM, which is depicted in Fig. 2. The selection of weights has been done empirically based on an expert's opinion over the importance and contribution of each metric to its parent metric. When generalizing on the SCM domain the MRG may be expanded, nevertheless it is omitted, due to space limitations.

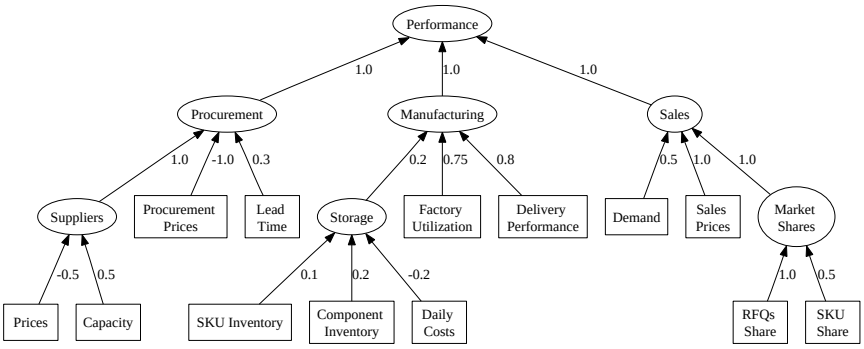

Fig. 2. A representational TAC SCM MRG. The `Performance` of the agent is divided into `Sales`, `Manufacturing` and `Procurement`, the three basic modules of a TAC SCM agent. Environmental impact is imported into the model with the nodes `Demand` and `Suppliers`

Employing the APE methodology we contribute towards robust and high performing TAC SCM agents by:

- Providing on-line agent performance measurements to aid mechanisms of on-line adaptivity of the agent.
- Identifying important changes in the strategies of other agents or the environment.
- Identifying the set of metrics and composite metrics that mostly affect agent performance.

- Reporting performance of agent modules or submodules of the system/agent.
- Provide a robust performance evaluator. Although forecasting the final agent performance is also plausible, we are mainly interested in comparing the performance of different implementations or modified versions of an agent.

5 Conclusions and Future work

In this paper, we have introduced the Metrics Representation Graph (MRG), a structure for organizing evaluation-specific knowledge, in a reusable fashion. We have also presented a full instantiation of an MRG for the domain of the TAC SCM competition. Future plans include indentifying the optimal ways of instantiating and evaluating the methodology described in Sections 3 and 4. Future research directions also include the inductive generation of graph weights. The process of inducing the values of the weights may be modeled as a supervised learning problem that can be solved through an error backpropagation algorithm, or a genetic algorithm that attempts to optimize the parameter space of the model.

6 Acknowledgements

This paper is part of the 03ED735 research project, implemented within the framework of the Reinforcement Programme of Human Research Manpower (PENED) and cofinanced by National and Community Funds (25% from the Greek Ministry of Development-General Secretariat of Research and Technology and 75% from E.U.-European Social Funding).

References

1. He, M., Jennings, N.R., Leung, H.F.: On agent-mediated electronic commerce. IEEE Transactions on Knowledge and Data Engineering **15**(4) (Jul/Aug 2003) 985–1003
2. Arunachalam, R., Sadeh, N.M.: The supply chain trading agent competition. Electronic Commerce Research and Applications **4** (2005) 66–84
3. Pardoe, D., Stone, P.: An autonomous agent for supply chain management. In Adomavicius, G., Gupta, A., eds.: Handbooks in Information Systems Series: Business Computing. Elsevier (2007)
4. He, M., Rogers, A., Luo, X., Jennings, N.R.: Designing a successful trading agent for supply chain management. In: Fifth International Joint Conference on Autonomous Agents and Multiagent Systems. (2006)
5. Sodomka, E., Collins, J., Gini., M.: Efficient statistical methods for evaluating trading agent performance. In: AAAI-07, Canada (July 2007)
6. Dimou, C., Symeonidis, A.L., Mitkas, P.A.: Data Mining and Agent Technology: a fruitful symbiosis. In: Soft Computing for Knowledge Discovery and Data Mining. Springer (2007)
7. Wellman, M.P., Jordan, P.R., Kiekintveld, C., Miller, J., Reeves, D.M.: Empirical game-theoretic analysis of the tac market games. In: AAMAS-06 Workshop on Game-Theoretic and Decision-Theoretic Agents, Hakodate, Japan (2006)