Combination Framework of BI solution & Multi-agent platform (CFBM) for multi-agent based simulations

Truong Minh Thai*, Frédéric Amblard*
Benoit Gaudou*

* UMR 5505 CNRS, Institut de Recherche en Informatique de Toulouse
Université Toulouse 1 Capitole
tmthai@cit.ctu.edu.vn, frederic.amblard@ut-capitole.fr, benoit.gaudou@ut-capitole.fr

Abstract. Integrated environmental modeling in general and specifically Multi-agent-based modeling and simulation approach are increasingly used in decision-support systems with, as a major consequence, to manipulate and generate a huge amount of data for their functioning (parametrization, use of real data in the simulation, ...). Therefore there is a need to manage efficiently these data being either used or generated by the simulation. Practically, existing generalist simulation platforms lack database access and analysis tools and simulation outputs are usually stored as text files or spreadsheets to be manipulated later by dedicated tools. In this paper, we propose a solution to handle simulation models data, i.e. their outputs as well as corresponding real data. We designed a conceptual framework based on a combination of two components, a Business Intelligence (BI) solution and a multi-agent platform. Such a framework aims at managing simulation models data throughout the lifespan of the simulation, from its execution and its coupling with real data to the generation of simulation results in order to use the simulation model as an effective decision-support system with what-if scenarios.
Keywords: Decision-Support System, Multi-Agent Simulation, BI, Data warehouse, OLAP.

1 Introduction

Integrated socio-environmental modeling in general and the multi-agent based simulation approach applied to socio-environmental systems in particular are increasingly used as decision-support systems in order to design, evaluate and plan public policies linked to the management of natural resources (Laniak et al., 2013). The main idea behind such approaches is to combine and couple information available from different sources and scientific fields (like water management, climate sciences, sociology, economics). Such information mainly take the form of empirical data gathered from the field and of simulated models regarding some aspects of the studied phenomenon (for instance the behavior of related actors from the field). This information is usually combined in an ad hoc manner, combining database access and management of text files, and simulation results are often stored as text files, spreadsheets or xml files. The basic statement we can make is that currently, if the design and simulation of models has benefited from the advances in computer science through the popularized use of simulation platforms (like Netlogo (Wilensky, 1999), GAMA (Taillandier et al., 2012)...), it is not the case for the management of data yet, which are still managed in an ad hoc manner, despite the advances in the management of huge datasets (data warehousing for instance). Such a statement is rather pessimistic if we consider recent tendencies toward the use of data-driven approaches in simulation aiming at using more and more data available from the field into simulated models.

Therefore, we stand that there is definitely a need for a robust data management solution of huge datasets in multi-agent based simulations and we will in this article, propose the first steps towards such a solution. This solution will combine two aspects. The first one deals with the status of data, as the proposed solution should be able to manage empirical data gathered from the studied phenomenon or system as well as simulated data produced by simulations considered as *in silico* experiments on the same system. The second aspect concerns the use of a Business Intelligence (BI) solution envisaged as a system of data warehouse and analysis tools. A data warehouse corresponds to a collection of data that supports decision-making processes (Inmon, 2005). Analysis tools may be data mining, statistical analysis, prediction analysis and so on. The features of a BI solution will help us to manage huge amount of historical data and make several analysis on such data.

In the following, we first present a state of the art of the links between these two systems (Section 2). We then present the global architecture of our combined framework (Section 3), before discussing its strengths and weaknesses (Section 4).

2 Related works and methodologies

Lot of works have been done to deal with huge amount of data and provide analysis tools in the field of Data Warehouse (DW) and BI tools. Many researches have been conducted on the use of DW and simulation to develop a decision support system or a prediction system. The combination of simulation tools and DW is widely increasingly used and applied in different areas. (Madeira et al., 2003) proposed a new approach dedicated to analyze and compare a large amount of output data from different experiments or similar experiments across different systems. They gathered data from raw data sources (text file or spreadsheet) into multidimensional database and use OLAP tools to analyze or compare them. (Sosnowski et al., 2007) proposed a data warehouse for collecting and analyzing simulation results. Although this is only an application of OLAP technologies to a special problem, namely the system dependability evaluation using fault injections into running programs, this work demonstrates that dimensional tables can store several hundreds of thousand records of simulation results. The multidimensional database of the simulation results can be analyzed, mined and enabled reporting by using standard OLAP tools. In (Vasilakis et al., 2008) and (Ehmke et al., 2011), the authors developed systems involving simulation models, multidimensional database and OLAP tools. These systems are called "decision support system" or "forecast system". Although these researches only solve specific problems, they demonstrated a potential in gathering and analyzing simulation results by using data warehouse and OLAP. (Mahboubi et al., 2010) presents a research on using multidimensional model to develop data warehouse in systems, coupling complex simulation models such as biological, meteorological and so on. These models are usually coupled models and they generate a huge amount of output data.

Most articles cited above focus on the application of data warehouse and OLAP technologies for collecting and analyzing output of simulations with huge amount of data. Although authors have successfully combined simulation models, data warehouse and OLAP technologies to solve concrete problems, they only focused on building specific simulation models and multidimensional schemas for solving specific problems. The state of the art demonstrates therefore the practical possibility and the usefulness of the combination of simulation, data warehouse and OLAP. It also shows the potential of a general framework that is, to our knowledge, not yet proposed in the literature.

Hence, the purpose of our research is to design a logical framework for the combination of both BI and MABS solutions. The framework can help us to exploit the useful features of BI solution and multi-agent platform to build a model, manage multiple models (especially their inputs/outputs) and analyze results of simulations. In particular, the framework should support a distributed and collaborative environment.

Our proposed Combination Framework of BI solution and Multi-agent platform, named CFBM for short, is detailed in the next session.

3 Combination Framework of BI solution & Multi-agent platform (CFBM)

3.1 Computer simulation system

Fishwick (1997) defined that computer simulation is the discipline of designing a model of an actual or theoretical physical system, executing the model on a digital computer, and analyzing the execution output. On the basis of Fishwick's definition, we can define a computer simulation system as below:

A Computer simulation system is a computation system with four components and the intercommunications between them:

- *Model design tool: a software environment that supports a modeling language, notations and user interface for modeling an actual or theoretical physical system.*
- *Model execution tool: a software environment that can run models.*
- *Execution analysis tool: a software environment that supports statistical analysis features for analysis of output data of models.*
- *Database tool: a software environment that supports appropriate database and database management features for overall components in the system.*

The components of a computer simulation system and their intercommunications are illustrated in Figure 1. From this figure, we have designed a conceptual framework for multi-agent based simulations. The framework is illustrated in figure 2. In this framework, we use a BI solution as a database tool, a multi-agent platform as model design tool and model execution tool. For execution analysis, we can either use OLAP analysis tool or use analysis features as an external plug-in of the multi-agent platform (for instance R scripts).

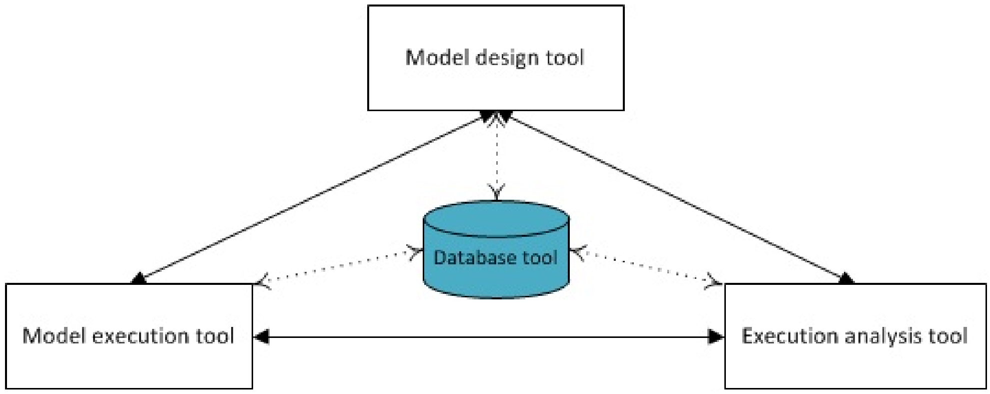

FIG. 1 – *Computer simulation system.*

3.2 CFBM architecture

In Figure 2, the CFBM is divided into three systems with seven layers. The function of each part is detailed below.

3.2.1 Simulation system

The simulation system is composed of a multi-agent platform with a relational database. This system helps to implement simulation models and to handle the various models and their input/output data. The multi-agent platform supports the model design and its execution.

The simulation is composed of the following parts. The **Simulation interface** is an user environment that helps the modeler to design and implement his models, execute them and visualize simulation results. **Multi-agent simulation models** are a set of multi-agent based models and can be considered either as alternative models of the same phenomenon or as different versions of the same model. They are used to simulate phenomena that the modeler aims at studying. A multi-agent simulation model reads input values from a database and store simulation results via SQL agents. The **SQL Agent** is a particular kind of agents which supports SQL features to query data from a relational database. As these agents are integrated into the simulation, they play an intermediary role between the simulation and the database layers.

The data source layer is composed of two relational databases. **Real data** is a database used to store empirical data gathered from the target system that are needed for simulation and analysis. This information can be used as input data for the simulation phase or as validation data for the analysis phase. **Simulation data** is a database used to manage the simulation models, simulation scenarios and output results of the simulation models.

In principle, the real data and the simulation data are separated, even if they are related to the "same objects" (one being from the real world, the other from the simulated world) and that both will be used to answer questions from the decision-maker in many cases. One of the important feedback to make to the decision-maker is related to the quality of the given answer, and a part of this quality can be represented in the rate of simulated versus empirical data used

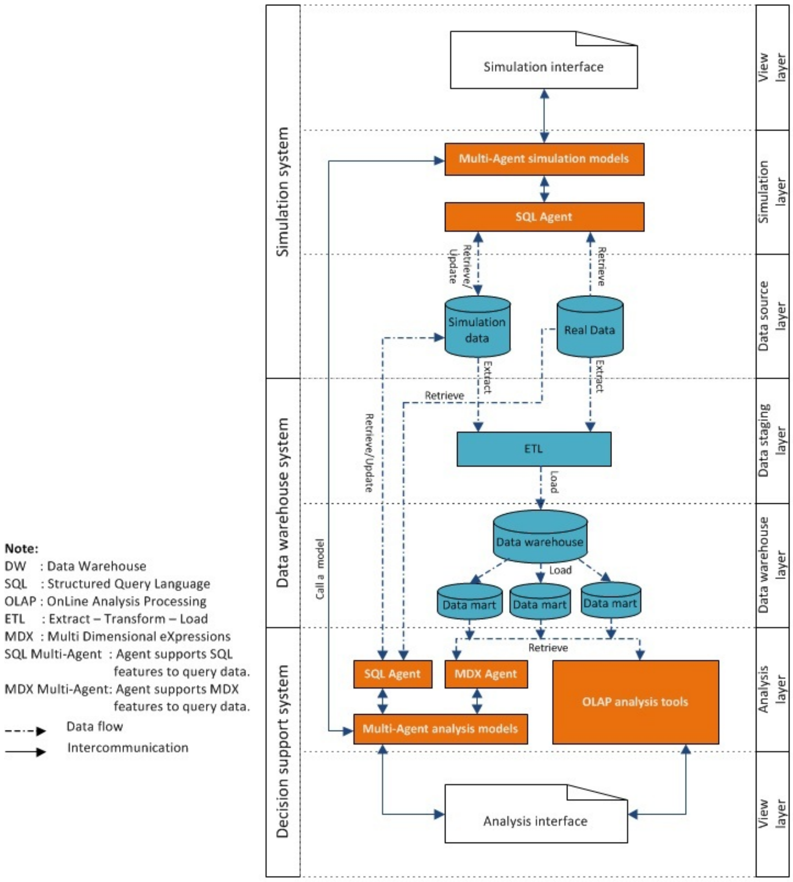

FIG. 2 – *Combination Framework of BI solution and Multi-agent platform (CFBM) architecture.*

to answer the question. In any case, these two data sources will be used to feed the second part of the framework, namely the **Data Warehouse system**.

3.2.2 Data warehouse system

The data warehouse system is used to store historical data about the actual system and simulation data. It is divided into three parts.

ETL is a set of processes with three responsibilities. First, ETL (Extract-Transform-Load) extracts data from the real and the simulation databases. Second, it transfers extracted data from these two sources into an appropriate data format. Finally, it loads transferred data into the data warehouse. **Data warehouse** is a relational database used to store historical data which are loaded from simulation system by ETL. A **Data mart** is a subset of data stored in the data warehouse. It is used to gather and store data for one or many analysis. We can create several data marts depending on our analysis requirements. Data marts are particularly useful to help users to improve the performance of analytic processes.

We propose to use a BI solution with OLAP technologies to implement the data warehouse system. OLAP is the most popular way to exploit information in a data warehouse (Golfarelli and Rizzi, 2009). OLAP technologies help experts to analyze multidimensional databases having a complex analysis requirement, with some common operators of OLAP such as roll-up, drill-down, slide and dice, analysts can aggregate data, navigate data to more details, slice data into specific datasets and dice data in different viewpoints easily. In addition, our framework has been designed modular enough to use other technologies instead of OLAP. In particular, OLAP could be replaced by an open-source alternative.

3.2.3 Decision support system

In our architecture, the decision support system part is a software environment supporting analysis, decision-making features and visualization of results. There exist many tools that could be used to implement such a decision support system. In our framework, we give the user the opportunity to use either existing OLAP analysis tools, or a multi-agent platform with analysis features or a combination of both of them.

The decision support system is thus composed of four parts. The **Analysis interface** is an user interface that is used to handle analysis models and visualize results. It should be generic enough to be used by the two decision support tools: the multi-agent and the OLAP analysis tools.

Multi-agent analysis models are a set of multi-agent-based analysis models. They are created based on analysis requirements and handled via analysis interface. One of the key points of our framework is the fact that multi-agent analysis models and multi-agent simulation models are implemented with the same modeling language in the same platform hence they can communicate easily with each others. The **MDX Agent** is a special kind of agent which supports MultiDimensionnal eXpressions (MDX) features to query data from a multidimensional database. Analysis model can access data from relational database or multidimensional database by using SQL agent (the same agent as in the simulation system) or MDX agent appropriately.

OLAP analysis tools is an analysis software that supports OLAP operators.

The key points of the CFBM architecture are that the model design, the model execution and the execution analysis functions are integrated into one multi-agent platform, all related data are managed by data warehouse, and analysis models and simulation models can interact with each others. They are very useful features for building a computer simulation system.

4 Discussion: Advantages and disadvantages of CFBM

CFBM is a conceptual framework that we have designed to manage interactions between multi-agent based simulations and large amounts of data. The framework has some advantages listed below.

CFBM is an open and modular architecture. As for the implementation of the CFBM, we can use any BI solution and multi-agent platform depending on which technology is the most adapted. We can choose an open source software or a commercial one. While choosing such a software, however we should consider important points such as the selected software adaptation with our requirements on model design, model execution and execution analysis, the integration seems feasible.

CFBM can be used in a distributed environment. Assuming many modelers and analyzers working together on the same project but being located in different places, we can setup a simulation system and an analysis system in each location and all the data from each location can be integrated and shared via a centralized data warehouse.

CFBM allows to handle a complex simulation system. In particular, we can build several simulation models to simulate the same phenomenon, conduct lot of simulations on each of them and compare simulation results of each model (e.g. to determine which one is better for which parameters value domain). In this case, it is very difficult for modelers to manage, analyze or compare output data of simulations if modelers do not have an appropriate tool. With help of SQL agents and relational data in simulation system part of CFBM, modelers can create a database to manage and store simulation models, scenarios and outputs of simulation models easily. In addition, ETL will load all the outputs of a simulation and appropriate empirical data into data warehouse then it also allows modeler to deal with several analyses to compare simulation results of a simulation in different scenarios as well as to validate simulation models with empirical data.

CFBM is a combination framework of BI solution and multi-agent platform, it supports simulation modeling, database management, database analysis and interaction environment between them hence CFBM is suitable to implement many kinds of system such as what-if simulation systems, prediction/forecast systems or decision support systems.

Although CFBM has many advantages, it still has some drawbacks such as:

- It is very hard to implement CFBM. Because CFBM is a mix system of different applications, multi-agent platform, BI solution and analysis tools, the integration of all these software components and expertise is a complex work, however similar complexity is actually present when working on integrated modelling.
- CFBM is not suitable for building a simple simulation system such as one or two models working with a small amount of data because it may take more time and workforce than other approaches.

5 Conclusion and future works

In this paper, we proposed a conceptual framework adapted to multi-agent-based simulations with high volume of data. Not only CFBM supports experts to model a phenomenon and to execute the models via a multi-agent based simulation platform, but our framework also helps experts to manage sets of models, input and output of models, to aggregate and analyze output data of models via data warehouse and OLAP analysis tools. The key features of CFBM are that it supplies four components (model design, model execution, execution analysis and database management) and it also assures communication in-between these components in a computer simulation system. The most important point of CFBM is the integration power of data warehouse, OLAP analysis tools and a multi-agent based platform that is useful to develop complex simulation systems such as what-if simulation system, prediction/forecast system or decision support system with a large amount of input/output data.

As future works, on the one hand we will choose an appropriate BI solution and a multi-agent based platform to implement our framework. The prototype of the architecture is being developed using the open-source GAMA platform (Taillandier et al., 2012). On the other hand, we will apply our framework on the management of simulation models and their input/output, analysis output simulations and evaluation models of two projects (the DREAM Project [1] and the MAELIA Project [2]). These two research projects aim at building integrated models, dealing thus with several models (and several versions of the same models) and huge quantities of data.

References

Ehmke, J. F., D. GroSShans, D. C. Mattfeld, and L. D. Smith (2011). Interactive analysis of discrete-event logistics systems with support of a data warehouse. *Computer Industry 62*(6), 578–586.

Fishwick, P. A. (1997). Computer simulation: Growth through extension. *Transactions of the Society for Computer Simulation International 14*(1), 13–23.

Golfarelli, M. and S. Rizzi (2009). *Data Warehouse Design: Modern Principles and Methodologies*, Chapter 1. Mc Graw Hill.

Inmon, W. H. (2005). *Building the Data Warehouse, 4th Edition* (4th Edition ed.). Wiley Publishing, Inc.

Laniak, G. F., A. E. Rizzoli, and A. Voinov (2013). Thematic issue on the future of integrated modeling science and technology. *Environmental Modelling & Software 39*(0), 1–2.

Madeira, H., J. Costa, and M. Vieira (2003). The olap and data warehousing approaches for analysis and sharing of results from dependability evaluation experiments. In *IEEE/IFIP International Conference on Dependable Systems and Networks (DSN'03)*, Los Alamitos, CA, USA, pp. 86. IEEE Computer Society.

Mahboubi, H., T. Faure, S. Bimonte, G. Deffuant, J. P. Chanet, and F. Pinet (2010). A multidimensional model for data warehouses of simulation results. *IJAEIS 1*(2), 1–19.

Sosnowski, J., P. Zygulski, and P. Gawkowski (2007). Developing data warehouse for simulation experiments. *Lecture Notes in Computer Science (LNCS) 4585*, 543–552.

Taillandier, P., A. Drogoul, D. A. Vo, and E. Amouroux (2012). GAMA: a simulation platform that integrates geographical information data, agent-based modeling and multi-scale control. In Springer (Ed.), *The 13th International Conference on Principles and Practices in Multi-Agent Systems (PRIMA)*, Volume 7057 of *Lecture Notes in Coputer Science*, India, pp. 242–258.

Vasilakis, C., E. El-Darzi, and P. Chountas (2008). A decision support system for measuring and modelling the multi-phase nature of patient flow in hospitals. *Studies in Computational Intelligence 109*, 201–217.

Wilensky, U. (1999). Netlogo. Technical report, Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

1. http://www.ctu.edu.vn/dream/
2. http://maelia1.wordpress.com/