# Simulation as a Sensor of Emergency Departments: Providing Data for Knowledge Discovery

Work-in-Progress Paper

Eva Bruballa, Manel Taboada
Tomàs Cerdà Computer Science School
Universitat Autònoma de Barcelona (UAB)
Barcelona, Spain
eva.bruballa@eug.es, manel.taboada@eug.es

Eduardo Cabrera, Dolores Rexachs, Emilio Luque
Computer Architecture and Operating Systems Department
Universitat Autònoma de Barcelona (UAB)
Barcelona, Spain
ecabrera@caos.uab.cat, dolores.rexachs@uab.es,
emilio.luque@uab.es

***Abstract*— Simulation of unusual or extreme situations of Hospital Emergency Departments (ED) makes it possible to get extra knowledge about the behavior of the system which could not be obtained in other way. There is no real data available of such situations, but simulation allows us to obtain this information. In this paper, we show how the data obtained by simulation of scenarios representing special real situations expand the real available data. That provides further information, which will allow us to reach more reliable models of real system behavior, avoiding extrapolation methods, which imply important errors, especially in nonlinear systems as ED. The objective pursued in this ongoing research is to extract the information contained in all this data to observe patterns and translate them into relationships and behavior models about variables which may influence the Hospital Emergency Department's performance and quality of service. A methodology to obtain new knowledge from intensive data, based on the use of the simulator as a sensor of the real system, and so as the main source of data, is proposed. As immediate future work, we intend to prove this methodology in a case of study which aim is to gain knowledge from a specific set of data, obtained through the simulation of a reduced set of scenarios of the real system.**

***Keywords—Agent-Based Modeling and Simulation (ABMS); Data-Intensive;Data Mining (DM); Decision Support Systems (DSS); Emergency Department (ED); Knowledge Discovery.***

## I. INTRODUCTION

### A. Simulation as a Decision Support System (DSS)

Hospital Emergency Departments (ED) are a primary care unit in healthcare systems and the main way of patient's admission to hospital. It is one of the most important components of the whole healthcare system and one of the most complex areas of a hospital.

The complexity of the ED service is due to its operation mode, which results from human actions and interactions between the different elements of which it is composed, and simulation based on an Agent-Based Model of the System (ABMS) becomes a powerful tool for its description. Simulation provides a better understanding of the system working way and of the activity of its elements, and it facilitates decision-making to establish strategies for its optimal operation [1][2]. Definitely, it is a way to achieve additional knowledge of the system, by developing inference processes on the variables of interest of the system in order to make predictions about the behavior of these variables under different conditions, based on information obtained from the generated data [3].

### B. Simulation as a sensor of the real system

This work in progress aims to obtain knowledge of the system, which could not be obtained without the use of simulation. In fact, the starting idea of this research is to understand the simulation as a "sensor" of the real system [4]. Such idea suggests the hypothesis of the ability to gain knowledge about the ED service behavior from the data provided by the simulation of any possible scenario. Then simulation allows us to obtain data from exceptional, ideal or extreme situations of the ED, such as extraordinary or unusual number of patients or abnormal number of staff (doctors or nurses). This information is very difficult to have in other way because it is impossible to prove these kinds of scenarios in the real system. Data obtained directly from the real system will be complemented with data generated by the simulator. This will allow us to obtain much more refined behavior models of the system, an extra knowledge which would not be possible to obtain without the simulation.

### C. An Agent-Based Model of the Emergency Deparment

Previous work carried out by our research group, High Performance Computing for Efficient Applications and Simulation (HPC4EAS), of the Universitat Autònoma de Barcelona (UAB), in collaboration with the ED Staff Team of the Hospital de Sabadell (one of the most important hospitals in Spain, which provides care service to an influence area of 500,000 people, and attends 160,000 patients per year in the ED), has led to the design and development of an agent-based-model of the ED. The model describes the ED's behavior from the actions and interactions between agents, and between them and the ED physical environment. The result is a simplified version of the ED which simulates patients with less acuity level (4 and 5 in the

 ISBN: 978-1-61208-371-1 

scale of priority and urgency that applies in Spanish hospitals), those causing saturation of the service most of the time. The implementation of the simulator has been done with NetLogo, an Agent-Based Simulation environment well suited for modeling complex systems, and it has been validated and verified, in this its first version [5][6].

The input parameters that characterize each different scenario in the simulation of the real system are the healthcare staff configuration, the number of incoming patients, the derivation percentage of patients and the period of time simulated. In addition, each simulation provides data of the Length of Stay (LoS) of all patients in all ED locations and the number of patients per hour and location [5][6].

### D. High Performance Computing (HPC)

There are a great number and variety of simulated agents, and different possible values for the input parameters in the simulator. This results in a large number of different possible scenarios to be simulated. There is no other limitation in the amount of data that can be generated than the computational time needed for the executions of the simulation. Using High Performance Computing (HPC) we can run the simulation for any possible scenario in reasonable computation time. Each execution will generate the corresponding data based on the designed model. Thus, the potential of HPC makes it possible to generate a very large number of data, store this data, process and analyze it to be transformed into information. The effective processing and analysis of this intensive data cannot be done with traditional statistical techniques, and the use of data mining techniques is required to discovering these yet unknown patterns in the data and complete the knowledge discovery cycle. The new knowledge of the system can be manifested through patterns, rules, associations, groups, constraints, trends, etc.

There are previous works in which data mining is applied in different processes within the Hospital Emergency Departments to obtain knowledge about them. But in all cases, the analyzed data comes directly from the real system. These other works can guide us, taking as a reference their objective of study and the algorithms that they have applied in each case. The prediction of the demand for health workers, the influence of different variables on the workload, the expected saturation of the service, the formation of bottle-necks in the connection with the hospital and the creation of a clinical recommender system, or diagnostic support, are the objectives of these referred works [7][8][9][10][11].

The paper is organized as follows: Section II presents the objectives and features of the research, and the proposed methodology; Section III describes a case study which is a reduced example of the proposed working method. Finally, section IV closes the paper with discussion and future work.

## II. Objectives and Proposed Methodology

The main purpose of this research is to obtain knowledge from the intensive data generated via simulation of the real system.

The specific objectives to be achieved are:

- To gain knowledge about the system behavior for any possible situation (e.g., epidemics, mass accidents, unexpected increase in the demand for service, etc.).
- To anticipate solutions for unusual situations: Provide prediction models of staff demand and other resources in the service, as a reference for making decisions in exceptional cases.

Our proposal of methodology is represented in Figure 1, which shows the process we will follow in our search from data to knowledge. From bottom to top, the data will be generated by the simulator, processed, transformed and integrated in a data warehouse, and finally analyzed and interpreted using data mining techniques to observe patterns and reach unknown models of system behavior.

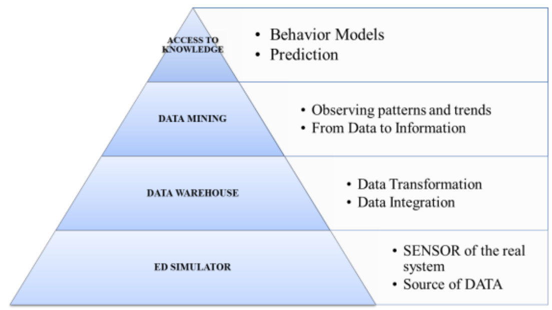


Figure 1. Phases of the proposed methodology

The phases of the proposed work system are described below in more detail.

### A. The simulator: Source of Data

Our ED simulator, already verified and validated, allows the generation of data concerning different possible situations, called scenarios, by assigning different values to the input parameters, either representing the more common conditions or other extreme or less likely situations, which are difficult to test in reality. This approach offers the possibility to obtain data for any possible scenario of service.

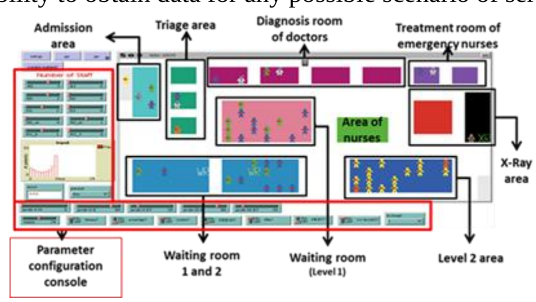


Figure 2. Screenshot with features of the simulator interface.

The present version of the simulator includes six primary areas: admission area, triage boxes, waiting rooms, diagnosis and treatment boxes, and x-ray area; and five different types of active agents: patients, admission staff, nurses, doctors and x-ray technicians (Figure 2). All incoming patients are triaged, but once they have been triaged, only patients identified at triage phase with acuity level 4 and 5 are served at the stage of diagnosis-treatment phase [12].

The input parameters that define each scenario are the staff configuration, the number of incoming patients (workload), the percentage of patient's derivation to other services, and the simulation period. Each possible configuration of the healthcare staff includes: admission personnel, triage nurses, doctors, emergency nurses and X-ray technicians. There is a minimum and maximum number of each type of staff, and two different levels of experience, low and high, labeled as junior and senior, respectively.

Furthermore, different values assigned to the random number generation seed, will affect the way incoming patients arrive into service. The distribution of patients' arrival changes, in each simulated hour, for a specific staff configuration, a given workload and a determined value for patients' derivation. This results in different iterations for the same scenario.

For each of the executions carried out, the simulator provides data concerning healthcare staff configuration, the time required by each of them in each location of the service, the cost of the configuration, and the Length of Stay (LoS) of each patient from their entry into the service until they leave the system, and specifically for each of the areas or places through which it passes. The number of patients per hour, per location, is also registered. Therefore, it is the assigned value for the rest of parameters defining the simulated scenario: the workload, the percentage of derivation for each type of patient, the simulation period and the random seed.

The data obtained directly from the real system is the first approach to obtain information. But, we can generate a much larger number of data using the simulator as a sensor of this real system, and thus, as data source, so that the reliability of the results will be much higher.

### B. Transforming Data: The Data Warehouse

After capturing the data, we can try to upgrade its quality, decoding, reformatting and merging it. Extract, Transform and Load (ETL) tools must help us to extract, scrub, transform and integrate all data in a proper way to be stored in a database, data mart or a data warehouse, to make its further analysis possible. Data partitioning, data combining or data summarization is the way to transform data and convert them from operational system format to data warehouse format.

### C. Observing Patterns: Data Mining

Stored data contains information that should be treated. We must deal with the analysis of large amounts of data, reason why statistical techniques and traditional methods are inadequate and difficult to apply.

Data mining techniques will be used to analyze such a large amount of data and try to extract the information behind them. Classification, association and clustering methods and algorithms, among others, are some of them. We have to choose which will be the best algorithm to find patterns or trends that will allow us, in turn, make inferences and prediction models through regressions between the observed variables.

### D. Acces to knowledge

Finally, this data analysis should lead us to knowledge discovery in terms of behavioral or prediction models. That might be helpful for decision-making in the operational management of the ED by the persons responsible for it.

## III. CASE STUDY

In this context and according to the agreed methodology, we will try to obtain knowledge from a set of data already available, obtained from the simulator execution for a reduced set of scenarios, carried out in a previous work within our research group [12].

In this case study, we deal with a set of scenarios characterized by the healthcare staff configuration and the number of incoming patients (workload). Any other input parameter is fixed for all the scenarios. The period of activivy simulated is the equivalent to one day (24 hours), the value for patient derivation percentage is zero, and the random seed is also fixed.

In the referred work [12], the total number of different staff configurations is 28,350. In addition, four different workload scenarios of total incoming patients, during the 24 hours of ED activity, are considered: up to 96, 216, 312, and 408 patients average per day.

In [12], it is described how exhaustive search among all configurations for a given workload was used to find out the optimal ED staff configuration that led to the minimum patient Length of Stay in the ED under a cost configuration constraint (the cost of staff configuration must be less or equal to €3,500) and the amount of human resources (ED Staff) available.

Moreover, in [12], it was implemented an alternative approach to reduce the computation time in simulation executions. This approach consisted of a combination of the Montecarlo method and K-means. After a random configuration execution by the Montecarlo method, the results of the K-means clustering method were that the scenarios were aggregated in some specific regions through the average LoS. Within the region containing the optimal solution, again an exhaustive search was carried out.

Now, we wonder about the possible relation between the set of staff configurations that are aggregated in the same cluster, as a result of this clustering method, applied to all the simulated scenarios. Do the points (configuration) of a specific region follow a specific law? Do they have something in common? Why have these regions and no others appeared? The goal is to observe patterns, similarities, any common feature among the staff configuration of all the scenarios of a same region. Then try to establish relations, and so, new knowledge of the system.

Around 25,000 configurations of the total 28,350 satisfy the considered cost restriction. These are the configurations that were executed for each workload. That is, about 100,000 scenarios were executed. These are the total of scenarios we will deal with for the verification of the proposed methodology, trying to find this possible relation between the aggregated configurations.

 ISBN: 978-1-61208-371-1 

The first step in our methodology process has been done, i.e., the data have been generated by the simulator. The second phase will be to transform and integrate these data into a data warehouse format.

This is the point we have reached in this work in progress. We know there is something hidden in the data from these 100,000 scenarios, and we think we have the right methodology to find it. Now, we have to choose the right algorithms and tools to attain this knowledge.

## IV. DISCUSSION AND FUTURE WORK

This ongoing research claims the benefits of simulation as source of data targeting a better understanding of the behavior of ED. Furthermore, the simulator, once tuned with the real system, is able to go beyond purely to replicate reality.

Epidemics, disasters and seasonal fluctuations on the demand of the healthcare service are real examples that could happen any time in the real system, and cause collapse, almost certainly, of the ED. Simulation is a way to obtain knowledge of these kinds of situations that can occur in reality, but cannot be tested in the real system before they happen.

The methodology proposed is supported by the use of high performance computing. HPC allows us to use the simulator as a sensor of the real system, to execute a large number of simulations, each one of them concerning an almost unlimited number of different scenarios, in order to obtain a large amount of data, which in many cases would not be available without the simulation.

Applying data mining techniques on such data will allow us to extract interesting patterns unknown before, thus, information about the real system, and finally knowledge about the behavior of the ED in such situations, which is useful for the decision making process.

As future work, a phase of experimentation with the simulator in its current version will be performed: Execute the simulation in all possible scenarios for different kinds of situations, to obtain intensive data to analyze and prove the proposed methodology to obtain some results.

We have to specify the tools and techniques for the storage and the processing of the data and find the suitable data mining algorithms to analyze it.

Simultaneously, in the same research line, we are working on the model of a new version of the simulator which includes patients with more urgent acuity level, that is, level 1, 2 and 3 in the mentioned scale of priority. Part of the future work is the implementation of this extended version of the simulator, which implies an enhancement that will greatly increase the number of possible scenarios to take into account in our future experimentation. The next experimental phase of the research will consist of executing new simulations with this extended version of the simulator, and again, trying to gain knowledge of the real system from the massive data generated by the improved sensor.

## ACKNOWLEDGMENT

This research has been supported by the MINECO (MICINN) Spain under contract TIN2011-24384.

## REFERENCES

[1] A. M. Mancilla, "Simulation: A tool for the study of real systems", Ingeniería y Desarrollo, Universidad del Norte, vol. 6, 1999, pp.104–112. Available from:
http://rcientificas.uninorte.edu.co/index.php/ingenieria/article/view/2226/1443 [Retrieved: July, 2014]

[2] J. Pavón, M. Arroyo, S. Hassan, and C. Sansores, “Simulation of social systems with software agents”, CMPI-2006, Actas del Campus Multidisciplinar en Percepcion e Inteligencia, vol. 1, 2006, pp. 389-400. Available from: http://samer.hassan.name/files/CMPI.pdf [Retrieved: July, 2014]

[3] L. R. Izquierdo, J.M. Galán, J.I. Santos, and R. Del Olmo, “Modeling complex systems using agent-based simulation and system dynamics”, “Modelado de sistemas complejos mediante simulación basada en agentes y mediante dinámica de sistemas”, Empiria: Revista de metodología de ciencias sociales, vol. 16, 2008, pp. 85–112.

[4] A. Choudhary, Northwestern University, USA, Keynote Speaker: “Big Data, exascale systems and knowledge discovery – The next frontier for HPC, Euro-Par Conference, August 2013. Available from: http://www.europar2013.org/upload/Dokumente/Euro-Par-2013-Keynote-Choudhary.pdf [Retrieved: July, 2014]

[5] M. Taboada, E. Cabrera, and E. Luque, “Modeling, simulation and optimization of resources management in hospital emergency departments using the agent-based approach”, Advances in Computational Modeling Research, 2013, pp.1–31.

[6] M. Taboada, E. Cabrera, F. Epelde, M. L. Iglesias, and E. Luque, “Using an agent-based simulation for predicting the effects of patients derivation policies in emergency departments”, Procedia Computer Science, vol. 18, ICCS 2013, pp. 641–650.

[7] C. C. Yang, W. T. Lin, H. M. Chen, and Y.H. Shi, “Improving scheduling of emergency physicians using data mining analysis”, Expert Systems with Applications, vol. 36, no. 2, 2009, pp.3378–3387.

[8] A. Ceglowski, L. Churilov, and J. Wassertheil, “Knowledge discovery through mining emergency department data”, Proceedings of the 38th Annual Hawaii International Conference on System Sciences, IEEE, 2005, pp. 142c–142c.

[9] A. Ceglowski, L. Churilov, and J. Wassertheil, “Combining data mining and discrete event simulation for a value-added view of a hospital emergency department”, Journal of the Operational Research Society, 2006, pp. 246–255.

[10] L. Grigull and W. M. Lechner, “Supporting diagnostic decisions using hybrid and complementary data mining applications: a pilot study in the pediatric emergency department”, Pediatric research, vol.71, no. 6, 2012, pp. 725–31.

[11] W. T. Lin, Y. C. Wu, J. S. Zheng, and M. Y. Chen, “Analysis by data mining in the emergency medicine triage database at a Taiwanese regional hospital”, Expert Systems with Applications, vol. 38, no. 9, 2011, pp. 11078–11084.

[12] E. Cabrera, M. Taboada, M. L. Iglesias, F. Epelde, and E. Luque, “Simulation optimization for healthcare emergency departments”, Procedia Computer Science, vol. 9, ICCS 2012, pp. 1464–1473.