IMPROVEMENTS IN THE DECISION MAKING IN SOFTWARE PROJECTS

I. Ramos Román , J. Riquelme Santos
Dpto. Lenguajes y Sistemas Informáticos. University of Sevilla.
Avda. Reina Mercedes s/n. Sevilla, Spain.
E-Mail: isabel.ramos@lsi.us.es, riquelme@lsi.us.es

J. Aroba Páez
Dpto. Ingeniería Electrónica, Sistemas Informáticos y Automática. University of Huelva.
Ctra. Huelva-La Rabida s/n. Huelva , Spain.
E-Mail : aroba@diesia.uhu.es

Keywords: Simulation, projects estimation, fuzzy logic, machine learning.

Abstract : The Simulators of Software Development Projects based on dynamic models have supposed a significant advance in front of the traditional techniques of estimate. These simulators enable to know the evolution of a project before, during and after the execution of the same one. But its use in the estimate of the project before beginning the execution, has been braked by the great number of attributes of the project that it is necessary to know previously. In this paper are presented the improvements that have been added to the simulator developed in our department to facilitate the use of them, and a new improvement obtained when using machine learning and fuzzy logic techniques with the databases generated by the simulator. In this last case, the project manager can know, in function of the decisions that he takes, the level of execution of the project objectives.

1. INTRODUCTION

In last years, the appearance of the dynamic models for Software Projects (SP) and of potent simulation environments (Stella, Vensim, iThink, Powersim, etc.), has facilitated the appearance of the denominated *Software Projects Simulator* (SPS), that provides the capability to simulate the behaviour of such projects. These simulators facilitates to the managers of projects to experiment with different management policies with a null cost, so that the decisions taken are as correct as possible.

Therefore, a SPS allows to carry out a:

1. A priori Analysis of the project: Simulation of the project, before beginning the development process, under different management policies and/or different initial estimates of resources and under the use of different development technologies.
2. Post-mortem analysis of the project: Simulation of projects that have concluded already, in those it's interesting to know how the obtained results could have been improved.
3. Project monitoring: Simulation of the project during the development process, to adapt the project estimations to the real evolution of the same one.

The SPS allows us to respond to questions of the type: "What will it happen if?" before beginning, "What is it happening?" during the execution and "What would happened if?" once it has concluded (Figure 1).

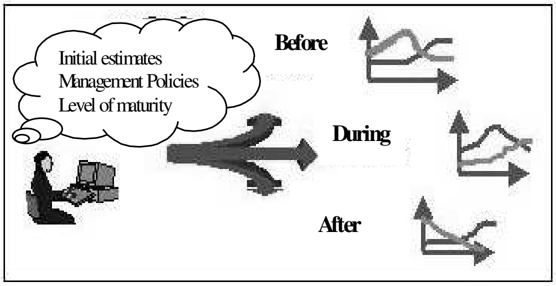


Figure 1: Possibilities of software projects simulators

The intuitive and natural idea contained in a dynamic model for SP, is that the evolution of the project and therefore the attainment of the goals, depends of: the initial estimate of the necessary resources to carry out the project, the management policies that are applied and certain aspects related with the own development organisation (level of maturity, experience, environmental aspects, etc).

The appearance, therefore, of the SPS it has meant a significant advance in the management and estimate of projects. Even this way, these simulators present an important inconvenience that have braked their use (to company level): the numbers of parameters (of the project and the development organisation) that should be known previously.

One of our first works was centered in the development of a SPS (Ramos 1998) that allows to simulate the behaviour of a project using half of attributes that the model that appears in (Abdel 1991). But we found an additional problem, common to the process that bears the decisions making in any project: the diversity and simultaneity of management policies, and factors that influence on the development process. That is, which is the appropriate combination of the attributes, to obtain some concrete objectives. The machine learning techniques allowed us to correct the previous problem (Ramos 1999). These techniques, applied to the databases that the SPS generates, provide to the project manager information about the management policies that he has to apply to get the objectives that previously, has established. In this paper, we present a step more in the way of improving the use of the SPS: the application of machine learning and fuzzy logic techniques.

In the following sections are described firstly the different contributions that have facilitated the use, not only academic but also managerial, of the SPS. Next, is presented the information that we can extract when conjugating the use of a SPS with machine learning and fuzzy logic techniques.

2. SOFTWARE PROJECT SIMULATORS

The basic variables that allow to know the behaviour of a dynamic system, and therefore the process of software development that we model by means of this system, are defined by means of differential equations. The system also includes a series of attributes that allow us to study the behaviour of the same one. These behaviors come given by the management policies that can be applied in the SP, so much those related with the environment of the project (number of tasks, time, cost, number of technicians, complexity of the product, etc.) as those related with the development organisation.

2.1 Initial situation

To simulate the behaviour of a SP by means of a SPS, the project manager would follow the following steps (Figure 2):

1. Definition of the concrete values that will take the attributes (parameters and functions) of the project.

2. Simulation of the model by means of the SPS to obtain the evolution of those variables that the manager wish to analyze

3. Analysis of the obtained results for these variables and check if this results coincide with the goals of the project (cost, quality, time, etc).

4. If the obtained results:

   - Coincide with the goals of the project, the parameters will be known, and therefore the management policies that will enable us to obtain the wanted results in the project.

   - Don't coincide with the goals of the project, it must be redefined again (step 1) and the manager must repeat the previous steps. Therefore, the project will be simulated so many times like be necessary until obtaining the wanted results.

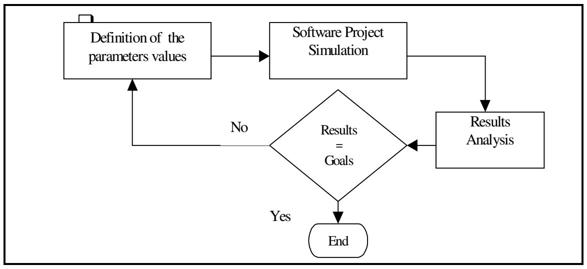


Figure 2 : Steps in the simulation of a Software Project using a SPS

Therefore, to diminish the number of times that the project will be simulated, the project manager must have an exhaustive level of knowledge of the concrete values that will take the model parameters; This can be quite difficult for two reasons: first, because the number of parameters that contains a dynamic model and the multiple combinations that can exist between them, and second, because frequently, the project manager does not know the concrete value that will take some parameters, although, he will have a clear idea of the interval of values where they will move. For example, the manager doesn't possibly know if the average daily manpower per staff in the project will be of 50% or of 24%, but he will know that it will be between a 20% and a 60%.

In the four steps described previously, only the step 2 is carried out in an automatic way. In the rest of the steps the project manager must intervene.

2.2 Improvements to the initial situation

The application of machine learning techniques, concretely the tool C4.5 (Quinlan 1993), allowed us to solve the previous problem. These techniques enable us to obtain management or decision rules[1]. The knowledge of these management rules helps in the decisions making to estimate the results wanted by the project manager. For the obtaining of management rules, we have conjugated the advantages that provides the systems that learn based on rules and the information that a dynamic model provides for SP (cost, delivery time, quality, etc). Also, the obtaining of management rules allows to the project director to analyze which are the most significant management policies, to get the objectives of the SP.

In the Figure 3 is shown the process used for the simulation of a SP and for the obtaining of these management rules, that allow to estimate these goals:

1. Definition of:
    - The intervals of values that can take the attributes of the project.
    - The project goals that the manager wants to obtain. For example, it is wanted that the final cost of the project doesn't overcome to the initial in more than 20% and that the final time of delivery doesn't overcome to the initial in more than 10%.
2. Simulation of the project by means of the SPS. For simulating the model, the parameters take values randomly chosen inside the interval of values defined previously.
3. Automatic generation of a database. In each simulation a record of the database is obtained; in these records are stored the values obtained for the parameters and the final values of the variables (obtained in the simulation) that interests us to analyse (cost, time, quality, etc.). Each register of the database represents a possible scenario of the project.
4. Application of machine learning techniques. Taking the generated database as input, the tool C4.5 learns examining the sample of resolved cases, and proposing a group of rules for the decision making. The C4.5 is a hierchical classifier, that provides semi-qualitative information.
5. Obtaining of management rules.

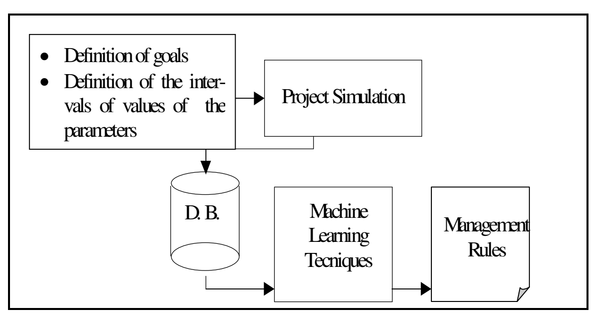


Figure 3: Steps in the simulation of a project using a SPS and machine learning techniques (Tool C4.5)

[1] We call management rule to a set of management policies (decisions) that the manager takes for carrying out the project goals.

In this case, the project manager should only carry out the first step, where the definition of the attributes is also facilitated regarding the step 1 of the previous section (figure 2); the rest of the described steps is carried out automatically

An example of management rule, obtained from the hierarchical tree that provides the C4.5, can be :

"If the average daily manpower per staff is greater than 50% and the medium time of the technicians' incorporation in the project it is smaller than 5 days and the dedication to formation it is then greater than 5% the wanted goals will be completed"

In summary, the use of a tool for the obtaining of management rules, allows to the project managers to analyze different situations, when they have to define the most appropriate management policies to optimize the final values of certain variables for separate or jointly.

Once obtained the management rules, the project manager will decide what rules or rules are the easiest of applying in function of the concrete project and of the software organisation in which he is working.

3. APPLICATION OF FUZZY LOGIC TECHNIQUES

The application of data mining techniques (clustering) and fuzzy logic, will allow us to add two new improvements in the use of a SPS:

1. Obtaining of qualitative management rules, that facilitates the interpretation of them.
2. The project manager doesn't have to define the results or wanted goals previously.

In the figure 4, are shown the steps to obtain this new type of rules. We can see (Step 1) how the project manager doesn't have to define the objectives of the project. In this case, an algorithm of fuzzy clustering that provides qualitative information, processes the database generated by the SPS.

This qualitative information that it's obtained from the quantitative data of the project's attributes, is a group of fuzzy management rules in graphic format,

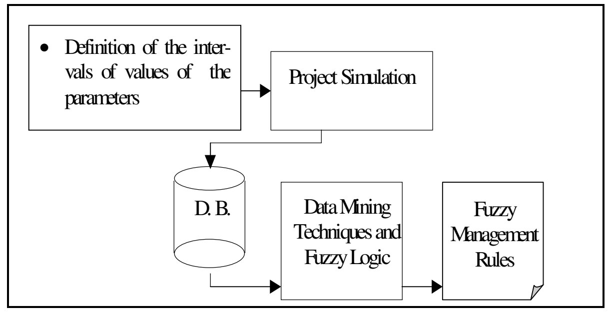


Figure 4 : Steps in the simulation of a project using a SPS and fuzzy clustering

that allow to the manager to interpret the project's data, in a quick and clear way.

Next, is described the used algorithm of fuzzy clustering, and are presented the obtained results.

3.1 Description of the used algorithm

The fuzzy clustering algorithm that will be applied to the quantitative data obtained by the simulation of the SP, to obtain the qualitative management rules, is based in the FCM algorithm (Fuzzy C-Means Method) (Bezdek 1981). Basically this algorithm tries to classify n elements $X_k \in X$ with $1\leq k\leq n$, with p features each one, that's to say, $X \subset \Re^p$, in C fuzzy clusters, and assigning a membership function μ_i:

$\mu\_i\_k : X \rightarrow [0,1]\ \forall i\ \ 1\leq i \leq C,\ \ \forall k\ 1\leq k \leq n$

The fuzzy clustering algorithm developed, follows the methodology proposed by Sugeno and Yasukawa (Sugeno 1993), that is applied to quantitative databases, with an only one-output parameter. Basically it consists on applying the fuzzy partition FCM to the output parameter. Once obtained a partition of the output space in fuzzy clusters, is carried out a projection of these clusters in the input space, obtaining a fuzzy set in $\Re^n$, that projected on each axis, assign to each characteristic a fuzzy set.

This methodology has been modified and improved, so that it can be applied to database with N output variables, also provide a more complete exit. The improvements that has been added to the original methodology are:

- The algorithm can work with quantitative databases, with N input parameters and M output parameters

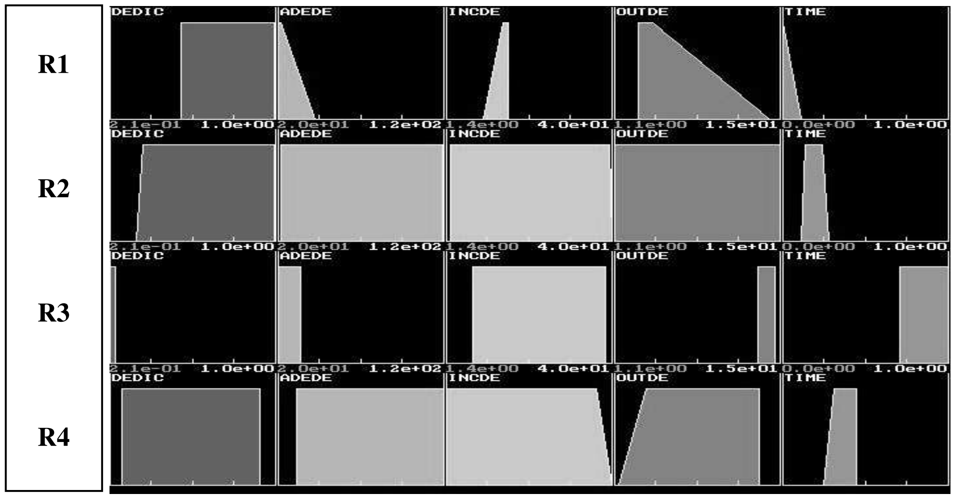


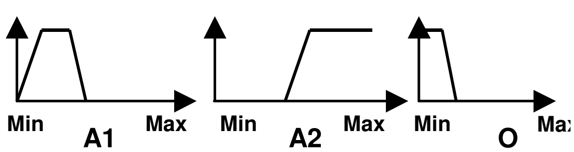
Figure 5 : Example of fuzzy managment rules with one output parameter (Time)

- The different variables can be pondered, assigning weights to calculate of the distances among points of the space that is being partitioned.

- Other algorithm, to obtain the trapezes of the graphic rules processes the obtained fuzzy sets.

- The output given by the original FCM algorithm has been improved, with a graphic interface, that shows the graphs of the management rules.

The fuzzy management rules obtained by applying to the database generated by the SPS, the fuzzy clustering algorithm, are represented in the following way:



In this fuzzy rule, the fuzzy set assigned to each attribute of the SP, is represented by a polyhedron. In the X-axis of each fuzzy set are represented the values of the parameter, and in the Y-axis the cluster membership values. This fuzzy management rule can be read as following:

If A1 is low *and* A2 is medium or high
Then O is very low.

3.2 Obtained results

In the Figure 5 is shown a set of fuzzy management rules obtained by applying the fuzzy clustering algorithm, described previously, to a quantitative database, obtained as result from the simulation of a project. In this figure are shown the fuzzy management rules that were obtained from a quantitative data series, with an only one-output parameter, in this case, the delivery time of the project (TIME). With this graphic representation of the rules, the project manager has a qualitative information of the SP, easy to interpret, that facilitate him the decisions making.

For example, in the third rule we can see that:

"The delivery time (TIME) takes high values when the average daily manpower per staff in the project (DEDIC) is very low, the adaptation delay of the new technicians (ADEDE) is low, the incorporation delay of the new technicians to the project (INCDE) is medium-high, and the exit delay of the unnecessary technicians in the project (OUTDE) is high".

The rest of rules indicate us the intervals of values that should take the parameters so that the development time be very low (R1), low (R2) and medium (R4). This kind of management rules can also be obtained for quantitative databases with N output variables simultaneously, so that the type of information that is provided to the project manager it

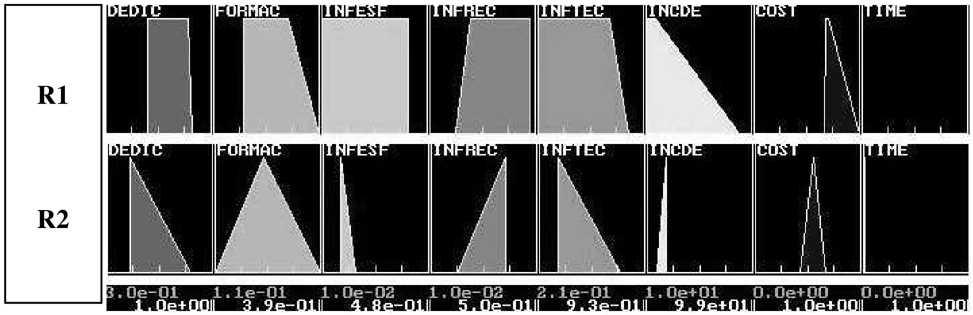


Figure 6 : Example of fuzzy mangement rules with two output parameters simultaneously (Cost and Time)

is much more complete, and it will facilitate him the decisions making.

As it has been commented previously, with the improvements that has been carried out to the described fuzzy clustering algorithm, we can use the number of input attributes and output variables that we need.

In the Figure 6 is shown a pair of fuzzy management rules obtained by applying the fuzzy clustering algorithm to a quantitative database, obtained as result from the simulation of a project with six attributes and two output variables: cost and time.

For example, the second rule (R2) indicates that:

"*The cost takes medium values and the time takes very low values simultaneously, when the average daily manpower per staff (DEDIC) is medium-low, the dedication to formation (FORMAC) is medium, the infra-estimation of the necessary effort (INFESF) is low, the infra-estimation of the project size (INFREC) is medium-low, the number of technicians that begins the project (INFTEC) is medium-high and the medium time of the technicians' incorporation in the project (INCDE) is low*".

4. CONCLUSIONS

In this work have been shown the advantages that present the SPS, based on dynamic models, in front of the traditional techniques of estimate (most of static character) and the inconveniences that, from our point of view, present the use of this simulators. These inconveniences are derived of the great number of attributes, so much of the project, as of the development organisation that previously should be known to use the simulator.

The combined use of a SPS and of machine learning techniques, as the C4.5, to facilitate the decisions making (by means of quantitative management rules) in the complex process of software development, allowed us to take an important step in our initial objective: facilitate the use of the SPS. Working in that way, at the moment we are adapting and applying fuzzy logic techniques to obtain qualitative information about the values of the attributes and the obtained results, of the project. In this case, the project manager leaving of the initial definition of the intervals of values that will take the attributes, on those that has bigger uncertainty, can know the level of execution of the project objectives.

ACKNOWLEDGMENTS

This work has been supported by the Spanish research agency CICYT under grant TIC99_0351.

REFERENCES

Abdel-Hamid T., Madnick S., 1991. "Software Project Dynamics: an integrated approach". Prentice-Hall.

Bezdek J.C., 1981. "Pattern Recognition with fuzzy objective function algorithm". New York : Plenum press.

Quinlan J.R., 1993. "C4.5 Programs for Machine Learning". Morgan Kaufmann Publ. San Francisco.

Ramos I., Ruiz M., 1998. "A Reduced Dynamic Model to Make Estimations in the Initial Stages of a

Software Development Project". In: Hawkings C et al (ed), INSPIRE III. Process Improvement through Training and Education, Pp.:172-185. London.

Ramos I., Riquelme J.C., 1999. "The Dynamic Models for Software Development Projects and the Machine Learning Techniques". International Conference on product focussed software process improvement. Pp.: 560-574. Oulu (Finland).

Sugeno M., Yasukawa T., 1993. "A Fuzzy-Logic Based approach to qualitative Modeling". IEEE Transactions on Fuzzy Systems, Vol.1.Pp.: 7-31.

Xiaohong Yuan, et all. 2000. “An application of fuzzy clustering to software quality prediction”. Symposium on Application-Specific Systems and Software Engineering Technology, Richardson, Texas USA. IEEE Computer Society. In press.

Taghi M. Khoshgoftaar, et all. 1999. ” Data mining for predictiors of software quality”. International Journal of Software Engineering and Knowledge Engineering. Forthcoming