Machine Learning applied to manage effort estimation and requirements in software projects*

Víctor Pérez-Piqueras[0000-0002-2305-5755]

Departamento de Sistemas Informáticos, Escuela Superior de Ingeniería Informática, Universidad de Castilla-La Mancha, Spain
victor.perezpiqueras@uclm.es

Abstract. Incremental complexity in software project management makes planning and decision making increasingly slow and difficult. Advances in Artificial Intelligence are shown as a possible solution to these management problems. In this way, it is intended to use Machine Learning algorithms applied to the field of software project management, in order to automate tasks and processes and support decision-making.

Keywords: Project management · machine learning · effort estimation · next release problem

1 Introduction to the research

As the years pass, software products and developments get more and more complex, requiring not only more effort in design and functionality, but also in planning and management. This implies that as the software gets more complex, so does its management. Thus, it has become necessary to create tools that help and support the software project management, in order to reduce its complexity, speed up the development process and support decision making. This line of research appears to solve many of the most common problems that arise in the field of software project management. In this research, two main problems have been tackled using Machine Learning algorithms:

1.1 Requirements prioritization

This problem consists of deciding the most important requirements to work on next, and it is also known as the "Next Release Problem" (NRP) [2]. It is a difficult and manual task that is performed many times throughout a project. To solve this problem satisfactorily, a certain balance must be found between the set of selected requirements, customer and stakeholder demands, and available

* This research statement and its initial results have been partially funded by the Junta de Comunidades de Castilla-La Mancha through the project SB-PLY/17/180501/000493.

resources.

Some of the algorithms that have been used to solve this problems are evolutionary algorithms (genetic, NSGA-II) [6], greedy algorithms (GRASP) [3] and Ant Colony [7, 8].

1.2 Effort estimation of requirements

Requirements are created to define specific needs that must be addressed in the software product being developed. To ease the project planning, requirements are often given an estimation of the effort that must be done to complete them. In plan-driven projects, this estimation is made once at the beginning of the project. But in projects applying value-driven methodologies, estimation of requirements is done iteratively many times during the projects life. Moreover, the requirements estimation is made by the development team, giving each requirement an expert value relative to the rest of the requirements. This repetitive, manual and experience requiring task is feasible to be solved by Machine Learning algorithms.

Machine learning models [5, 4, 1] have been applied to estimate the size of user stories. Neural networks offer good results compared to the rest; in all cases, the relevant information used is the text of the product backlog items. When working with textual information, the development of rule-based algorithms, descriptive or based on probabilistic graphic models could offer a great capacity to explain the prediction returned.

2 Overview of the Research Design and General Methodological Approach

This research will focus on two specific aims that will tackle the problems described previously:

- **A1. Search-Based Software Engineering applied to Next Release Problem.** Resolution of the NRP problem, using evolutionary and greedy algorithms. Special emphasis on the development of single and multivariate EDAs (Estimation of Distribution Algorithms), since they explicitly deal with the relationships between the variables. In addition, some of the models that can be created contain a structure that allows relationships between the elements of the problem to be identified. This is of special interest since, up to the author's knowledge, no EDA has been applied to the NRP problem yet. New forms of representation of the problem will be developed in addition to the classic one of requirements vectors, associated cost and list of stakeholders: Dependencies between logical and physical requirements, return on investment, forced order, uncertainty in importance of requirements, quality of requirements, etc.

– **A2. Supervised classification applied to effort estimation of requirements.** Neural networks will be used to generate automatic estimation of the effort of requirements. These neural networks will use textual information, such as requirement title and description to predict the estimation.

3 Expected Deliverables

– Reports for all experiments regarding A1 and A2.
– User interface and code shared with research community for algorithms which provided best results.
– Publication in journals from the JCR ranking. Candidates are shown in Table 1.

Table 1: Candidates for publications.

Journal	Category	Impact Factor
Knowledge-Based Systems	Computer Science, Artificial Intelligence	5.921
Applied Soft Computing	Computer Science, Artificial Intelligence	5.472
Automated Software Engineering	Computer Science, Software Engineering	1.857
Journal of Systems and Software	Computer Science, Information Systems and Software	2.450
Empirical Software Engineering	Computer Science, Software	3,156
Expert Systems with Applications	Computer Science, Artificial Intelligence	5,452
Information and Software Technology	Computer Science, Computer Science Applications, Information Systems, Software	2,726
Symposium on Search Based Software Engineering	Conference	
International Symposium on Empirical Software Engineering and Measurement	Conference	
International Conference on Requirements Engineering	Conference	

4 Schedule of Activities

The schedule of activities is shown in Figure 1.

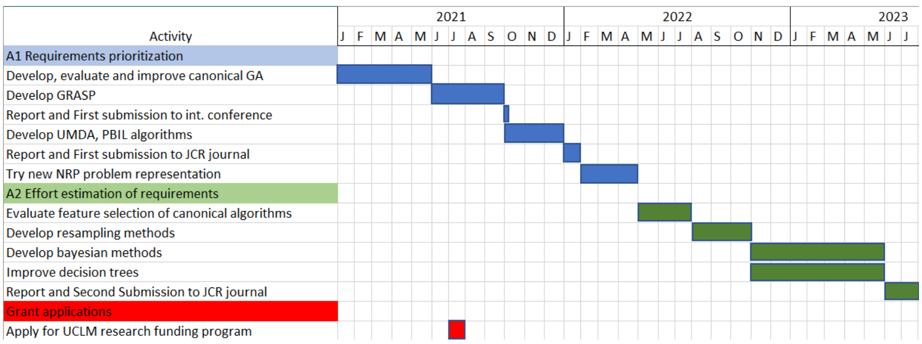


Fig. 1: GANTT diagram of activities.

5 Preliminary results

The specific aim A1 of the research has already started being tackled. To solve the requirements prioritization problem or NRP, an experimentation model has been designed. The experiments launched to evaluate each algorithm execute multiple times different combinations of the parameters, calculating final metrics that will be used to evaluate the performance of each one of the algorithms.

In the research, three algorithms have been designed to solve the NRP: a genetic algorithm designed by the authors, that stores a set of non dominated solution and uses a mono-objective metric as fitness function; the NSGA-II (Non Sorted Genetic Algorithm), a multi-objective algorithm frequently used to solve the NRP; and GRASP (Greedy Randomized Adaptive Search Procedure) algorithms.

The results obtained by the first two algorithms have already been analyzed, selecting the best configurations and comparing them to find their weaknesses and strengths (see Figures 2 and 3). The results were generated by applying the algorithms to two datasets [3] that represent instances of the NRP.

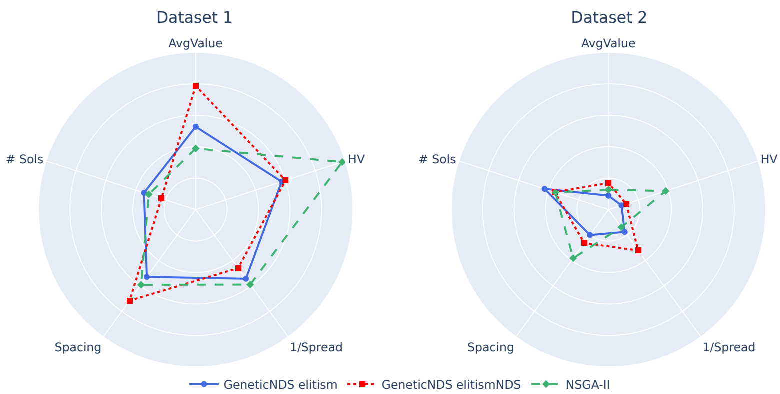
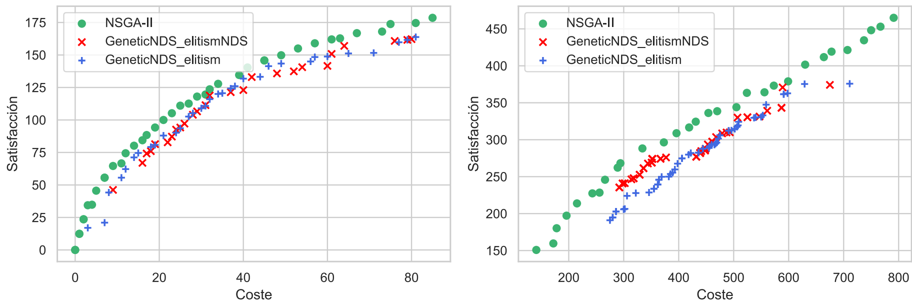


Fig. 2: Kiviat graph of metric comparison between algorithms.



(a) Results for Dataset 1

(b) Results for Dataset 2

Fig. 3: Optimal Pareto fronts of the best configurations of the algorithms analyzed.

References

1. Abrahamsson, P., Fronza, I., Moser, R., Vlasenko, J., Pedrycz, W.: Predicting development effort from user stories. In: International Symposium on Empirical Software Engineering and Measurement. pp. 400–403 (2011)
2. Bagnall, A.J., Rayward-Smith, V.J., Whittley, I.M.: The next release problem. Information and Software Technology **43**(14), 883–890 (dec 2001)
3. Chaves-González, J.M., Pérez-Toledano, M.A., Navasa, A.: Software requirement optimization using a multiobjective swarm intelligence evolutionary algorithm. Knowledge-Based Systems **83**(1), 105–115 (jul 2015)
4. Choetkiertikul, M., Dam, H.K., Tran, T., Pham, T., Ghose, A., Menzies, T.: A Deep Learning Model for Estimating Story Points. IEEE Transactions on Software Engineering **45**(7), 637–656 (2019)
5. Dantas, Emanuel; Perkusich, Mirko; Dilorenzo, Ednaldo; Perkusich, A.: Effort Estimation in Agile Software Development: an Updated Review. In: International Conference on Software Engineering and Knowledge Engineering (2018)
6. Durillo, J.J., Zhang, Y., Alba, E., Harman, M., Nebro, A.J.: A study of the biobjective next release problem. Empirical Software Engineering **16**(1), 29–60 (feb 2011), https://link.springer.com/article/10.1007/s10664-010-9147-3
7. Sagrado, J., Águila, I., Orellana, F.: Multi-objective ant colony optimization for requirements selection. Empirical Software Engineering **20**, 577–610 (2015)
8. de Souza, J.T., Maia, C.L.B., Ferreira, T.d.N., do Carmo, R.A.F., Brasil, M.M.A.: An Ant Colony Optimization Approach to the Software Release Planning with Dependent Requirements. In: Cohen, M.B., Ó Cinnéide, M. (eds.) Search Based Software Engineering. pp. 142–157. Springer Berlin Heidelberg, Berlin, Heidelberg (2011)