Goals at Risk? Machine Learning at Support of Early Assessment

Paolo Avesani, Anna Perini, Alberto Siena, and Angelo Susi
Fondazione Bruno Kessler
Trento-Povo, I-38123 Italy
{avesani,perini,siena,susi}@fbk.eu

***Abstract*—A relevant activity in the requirements engineering process consists in the identification, assessment and management of potential risks, which can prevent the system-to-be from meeting stakeholder needs. However, risk analysis techniques are often time- and resource- consuming activities, which may introduce in the requirements engineering process a significant overhead. To overcome this problem, we aim at supporting risk management activity in a semi-automated way, merging the capability to exploit existing risk-related information potentially present in a given organisation, with an automated ranking of the goals with respect to the level of risk the decision-maker estimates for them. In particular, this paper proposes an approach to address the general problem of risk decision-making, which combines knowledge about risks assessment techniques and Machine Learning to enable an active intervention of human evaluators in the decision process, learning from their feedback and integrating it with the organisational knowledge. The long term objective is that of improving the capacity of an organisation to be aware and to manage risks, by introducing new techniques in the field of risk management that are able to interactively and continuously extract useful knowledge from the organisation domain and from the decision-maker expertise.**

I. Introduction

In order to improve the quality of requirements, analysts need to understand the environment in which the system-to-be is going to operate, and the risks that can affect its capability to support stakeholder goals. Only with this knowledge, appropriate choices can be made and actions be taken to deal with problems that can arise. However, identification of risks specific to an enterprise and impacting its objectives and assets requires a huge effort for the management staff and for the analysts as well to analyse and synthesise different sources of knowledge and their own judgement. Therefore, while in some specific, highly critical contexts (e.g., aerospatial, railways, ...) the risk analysis activity is cost-effective, in most cases this activity remains unfulfilled, because too costly in terms of money, time, etc. Lacking in performing this activity may trigger the possibility of project failure [1].

To cope with this problem, we have conceived an approach to approximate an analytical risk analysis activity with the prioritisation of the different categories of risks potentially impacting the goals and qualities of the system, such as security, maintainability or legal compliance risks [2]. The sources of knowledge to identify risks can be of different nature as for example the event logs to extract risky situations, the expert opinions on what may go wrong in the enterprise, organisational documents describing criticality in the business processes, the simulation of business processes to extract a list of undesirable results, or the learning from experience of the components of the organisation. Some of these sources are able to produce only a list of risks, while other may produce some more or less accurate approximation of the frequency of the risk events actually happening so inducing a rank of the risk events in terms of their importance and criticality for the organisation. This aspect is particularly critical in some specific decision-making circumstances, such as those emerged in RISCOSS [3], [2], a project devoted to the specification of methods for the identification and management of risks in Open Source Software adoption. In this case a major problem has been that of supporting decisions to plan risk management and mitigation strategies to protect relevant goals, qualities and assets of the software system under production starting from the early phases of the development process. A relevant decision problem here is that of identifying the importance of the impact of different categories of risks on the goals of the organisation, such as software security and quality risks affecting not only the technological objectives of the organisation (e.g., the maintenance of good code quality) but also its strategic and business goals (e.g., maintain the capacity of selling products). In this case the organisations relay on different roles able to assess the different categories of risks. Moreover, organisations maintain documents and reports describing previous risk evaluations at the different levels of the organisational structure to be exploited in subsequent risk evaluation sessions.

In general terms the problem we aim at addressing can be stated as in the following. Given a set of organisational goals that are affected by different possible risks, the idea is that of finding a ranking of the goals with respect to the level of risk the decision-maker estimates for them, also exploiting and learning from risk related information already present in the documents the organisation maintains to complement the knowledge of the evaluator.

In this paper we describe a research activity we recently started, focusing on the general problem of risk decision-making, that aims at combining knowledge about risks assessment techniques and Machine Learning approaches to promote the active intervention of human evaluators in the decision

978-1-4673-6905-3/15/$31.00 © 2015 IEEE RE 2015, Ottawa, ON, Canada
RE:Next! Track

process and its integration with the organisational knowledge. The long term objective is that of improving the capacity of an organisation and of its management to be aware and to manage risks, by introducing new techniques in the field of risk management that are able to interactively and continuously extract useful knowledge from the organisation domain and from the decision-maker expertise.

In Section II the paper presents the generic definition of the problem, the background and the proposed approach. The work done so far and the next steps will be presented in Sections III and IV, finally, some conclusions are drawn in Section V.

II. Problem and Proposed Approach

In this section we report on the specific risk assessment problem we face with and the techniques we aim at exploiting for this purpose. In particular, the main ingredients of interest in our case are the concepts of risks and their influence on goals and the Machine Learning techniques we use to support the decision making related to the identification of the risk exposure of goals.

A. Formalisation of the Problem

Given a set of goals of the organisation that are affected by different risks we aim at finding a ranking of the goals with respect to their exposure to the set of risks based on the evaluation of decision-makers. So, given a set of goals $G = \{g_1, \ldots, g_n\}$ the problem is that of identifying a ranking of the goals in G based on the evaluation of how much those goals are more exposed to risk.

Together with the above elements, we also have a set of predefined ranking criteria ordering goals. They are a description on how much the different goals of the organisation are exposed to different kind of risks and are specified using knowledge already available in the organisation (for example, in form of documents related to previous risk assessments). In our formalisation these criteria are expressed as a set of m partial ranking functions $risk : G \to \mathbb{R} \cup \perp$, that assign a real or unknown ($\perp$) value to the goals in G.

B. Machine Learning Approach

The candidate Machine Learning approach is the CBRank method. It rests on a framework, introduced in [4], [5], and exploited in other works on requirements and testing [6], [7]. CBRank supports an *ex-post* decision-making process for ranking a set of alternatives, such as software requirements, on the bases of a predefined criteria. The use of machine learning techniques allows the approach to combine different rankings of the set of alternatives already known in the organisational domain and the priorities on the alternatives expressed by the decision makers in form of pairwise comparisons. Moreover, the approach exploits the domain knowledge to reduce the amount of information elicited from the decision-makers for achieving a ranking of a given quality degree; in fact, the method is able to identify priorities that are already embedded in the available domain knowledge and to reuse it in the calculation of the final ranking.

A characterising aspect of CBRank prioritization process is that of actively involving the human decision-maker by interleaving human activities with machine computation as described in Figure 1, where three steps are represented as rounded corner rectangles. The input and output artefacts, connected with dashed arrows to the basic activities, are: the set of `Goals` (G), the decision maker's `Priorities on Goals` on the exposure of goals to risks, the set of `Risk-Goals Ranking Functions` ($risk_i : G \to \mathbb{R} \cup \perp$), encoding the domain knowledge about the importance of the exposure of a given goal to a particular risk and the `Final Risk based Rank` ($risk_{final} : G \to \mathbb{R}$). Additional artifacts are produced by internal activities of the process's steps, namely: the set of `Sampled Goal Pairs` that is the set of goal pairs for which the end decision maker assessment is unknown and that have been selected for the following priority elicitation; the set of `Ordered Goal Pairs` that is a set of goal pairs ordered by the decision makers with respect to the exposure to risk. The process is based on the iteration of the three steps.

- In the `Pair Sampling` step an automated procedure selects from the set of `Goals` a set of `Sampled Goals Pairs` whose relative preference is unknown, according to an acquisition policy, which may take into account information on the currently available rankings, in order to maximise the usefulness of the information that will be elicited in the following Priority Elicitation step. An example, of sampling policy could be that of giving privilege to pairs of goals that appear in the first part of the rankings in order to improve accuracy there.
- The `Priority Elicitation` steps takes the collection of `Sampled Goals Pairs` produced by the Pair Sampling step as input, and produces in output a set of `Ordered Goals Pairs` on the basis of the `Priorities on Goals` expressed by the decision maker.
- In the `Priority Learning` step, given a partial elicitation of the stakeholder priority and a set of `Risk-Goals Ranking Functions`, the learning algorithm produces an approximation of the unknown preferences and then the correspondent `Risk based Goals Rank` for the goals thanks to the available functions. Several possible algorithms can be applied in this step. Among them we experimented the use of a method based on the use of a linear combination of several binary classifiers each one able to divide the space of alternatives in two subsets in which every element of a subset are ranked before those in the other subset [8]. The combination of these classifiers will produce a (partial or total) rank of the alternatives; the objective is that of minimising the difference between the produced rank, the ranks of the decision-maker and the rankings derived by the domain knowledge (the `Risk-Goals Ranking Functions`).

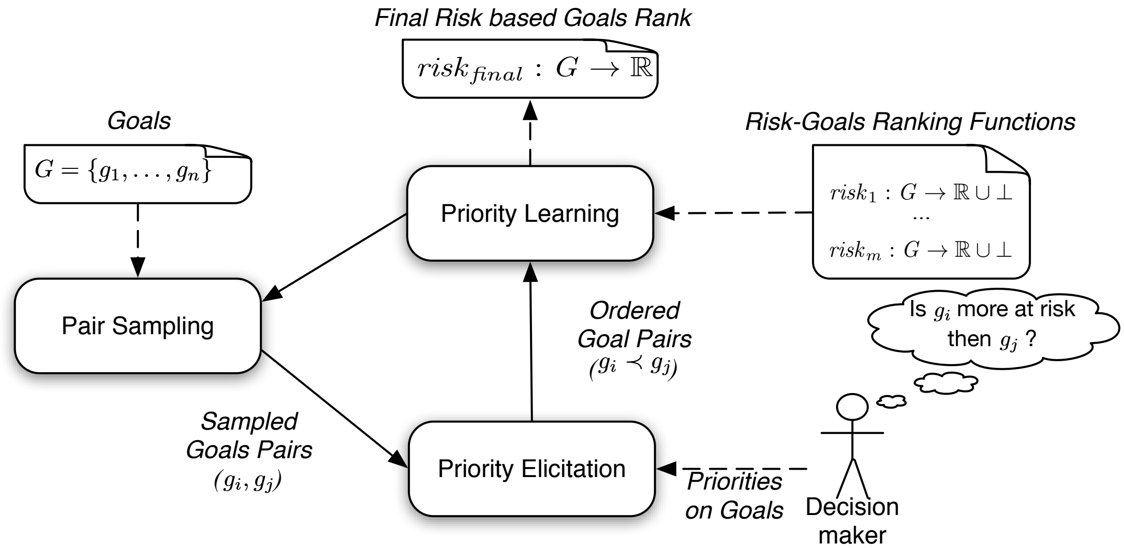

Fig. 1. The three steps of the risk based goal prioritization process in using the CBRank approach

If the result of the learning step is considered accurate enough (or time to input preferences runs out), the iterations stop and the process gives the last approximated rank (`Final Risk based Goals Rank`) as output.

III. Initial Results

In the initial phase of the work we focused on the preparation of a first definition of the problem mainly based on the identification of the needs of medium and large organisations we collected during RISCOSS [3]. In this case a major emerging problem has been that of optimising risk-related decisions to plan the risk mitigation strategies.

To this aim we identified CBRank as a suitable Machine Learning technique to be used to extract knowledge from the decision-maker; its strength stays in the possibility to combine this knowledge with the knowledge already available in the domain. As already observed, this is a quite usual situation in medium and large organisations where knowledge about risks is spread through several documents and risk related decisions are taken by different organisational roles having their specific competencies.

IV. Next Steps in Research and Application

In the next future we see several areas of development of this research line. First, we envisage a refinement of the CBRank technique in order to better adapt it to the particular problem of risk decision-making. Specifically, we aim at increasing the capability of the method to extract knowledge from the evaluators and to mix it with the other sources. An interesting problem here is that of considering multi decision-makers working in parallel or in collaboration on sets of risks in which they have specific competencies allowing then the method to produce a syntheses of the different perspectives (see also approaches for negotiation and decision-making in requirements engineering such as those described in [9]). This is actually one of the most relevant problems reported by the organisations participating in the RISCOSS project and, more generally, a relevant problem in project management practices as recognised in [1].

We also aim at studying the possibility to elicit the existence of risks not envisaged *a-priori*, that are tacit in the mind of the decision makers, and that can be discovered tanks to the pairwise comparison process. This kind of comparison, in fact, tends to stimulate a reflection on the why two cases (in our case goals) are different not only considering the quantitative difference of their exposure but also considering the different nature of the risks they are exposed to. In this case the possible strong difference between organisational knowledge and decision-maker evaluation can be a symptom of the presence of tacit knowledge that has been used by the human during the evaluation and that can be possibly explicated and argumented to extract its rational.

Finally, an important aspect is that of proposing a risk management process that can support the integration of the new techniques and that can be smoothly joined with the usual risk management practices already exploited in organisations.

V. Conclusion

The paper described a research work about risk assessment we recently started thanks to the identification of decision-making issues related to the management of risks in Open Source Software adoption in medium and large organisations. The work aims at introducing Machine Learning techniques in the solution of such problems in order to promote a deeper connection between available organisational knowledge and expert judgements about risks. In the next years we see two concrete action lines: the adaptation of generic Machine Learning techniques to the specific problem of risk management and the development of multi decision-maker collaborative approaches supported by these techniques.

ACKNOWLEDGMENT

This work is a result of the RISCOSS project, funded by the EC 7th Framework Programme FP7/2007-2013 under the agreement number 318249.

REFERENCES

[1] P. M. Institute, Ed., *Practice Standard for Project Risk Management*. PMI, 2009.

[2] A. Siena, M. Morandini, and A. Susi, "Modelling risks in open source software component selection," in *Conceptual Modeling - 33rd International Conference, ER 2014, Atlanta, GA, USA, October 27-29, 2014. Proceedings*, ser. Lecture Notes in Computer Science, E. S. K. Yu, G. Dobbie, M. Jarke, and S. Purao, Eds., vol. 8824. Springer, 2014, pp. 335–348. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-12206-9_28

[3] X. Franch, A. Susi, M. C. Annosi, C. P. Ayala, R. Glott, D. Gross, R. S. Kenett, F. Mancinelli, P. Ramsamy, C. Thomas, D. Ameller, S. Bannier, N. Bergida, Y. Blumenfeld, O. Bouzereau, D. Costal, M. Dominguez, K. Haaland, L. López, M. Morandini, and A. Siena, "Managing risk in open source software adoption," in *ICSOFT 2013 - Proceedings of the 8th International Joint Conference on Software Technologies, Reykjavík, Iceland, 29-31 July, 2013*, J. Cordeiro, D. A. Marca, and M. van Sinderen, Eds. SciTePress, 2013, pp. 258–264. [Online]. Available: http://dx.doi.org/10.5220/0004592802580264

[4] P. Avesani, S. Ferrari, and A. Susi, "Case-Based Ranking for Decision Support Systems," in *Proceedings of Case-Based Reasoning Research and Development, 5th International Conference on Case-Based Reasoning, ICCBR 2003*, ser. LNCS, no. 2689. Trondheim, Norway: Springer-Verlag, 2003, pp. 35 – 49.

[5] A. Perini, A. Susi, and P. Avesani, "A machine learning approach to software requirements prioritization," *IEEE Trans. Software Eng.*, vol. 39, no. 4, pp. 445–461, 2013. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/TSE.2012.52

[6] P. Avesani, C. Bazzanella, A. Perini, and A. Susi, "Facing scalability issues in requirements prioritization with machine learning techniques," in *Proceedings of 13th IEEE International Conference on Requirements Engineering (RE 2005)*. Paris, France: IEEE Computer Society, August, September 2005, pp. 297–306.

[7] P. Tonella, P. Avesani, and A. Susi, "Using the case-based ranking methodology for test case prioritization," in *Proceedings of 22nd IEEE International Conference on Software Maintenance (ICSM 2006)*, 2006, pp. 123–133.

[8] Y. Freund, R. D. Iyer, R. E. Schapire, and Y. Singer, "An efficient boosting algorithm for combining preferences," *Journal of Machine Learning Research*, vol. 4, pp. 933–969, 2003. [Online]. Available: http://www.jmlr.org/papers/v4/freund03a.html

[9] A. Finkelstein, M. Harman, S. A. Mansouri, J. Ren, and Y. Zhang, "A search based approach to fairness analysis in requirement assignments to aid negotiation, mediation and decision making," *Requir. Eng.*, vol. 14, no. 4, pp. 231–245, 2009.