Knowledge Extraction of the Behaviour of Software Developers by the Analysis of Time Recording Logs

Arturo Peralta, Francisco P. Romero, Jose A. Olivas, Macario Polo

Abstract—Software development project management has a poor reputation in terms of avoiding cost and schedule overruns. The cause of this situation is based on the feature of the software development process that is characterized by quickly growing complexity and change. Therefore, there are many uncertainties to define exactly the necessary time to complete a task according to the person’s performance. In this scenario, Soft-Computing techniques may offer new approaches with the aim of helping the participants of the project to manage their time, give priority to their activities and readjust the work to complete satisfactorily the project tasks. This work presents an automatic features extraction process with the aim of defining the elements involved in a software project. This knowledge is represented by means fuzzy sets and fuzzy prototypes. The source of data is the Personal Software Project time recording logs. A preliminary experiment illustrates the feasibility of this approach.

I. INTRODUCTION

Software project management requires gathering and processing an incessant stream of information and knowledge. New technologies, new customers and new knowledge change the software products and tools. Moreover, priorities and resources are both, in general, not well defined and they are constantly changing. Software processes are complex, interrelated, and often reflect large bodies of information. Small changes may affect many parts of a process description. If the project manager had planned the project around requirements that are changed, his planning would finish in scratch and earlier plans, decisions and relevant data are forgotten or neglected.

There are a lot of classic approaches applied to industrial activities or knowledge management. They are oriented more to the reuse of existing knowledge than the processing of incoming information, i.e., their purpose is directed more towards computer or industrial systems than human organisations. However, software development is a dynamic environment where some drawbacks are rooted at the personal level [1]. The “human factor” of software is added to the problem of task definition, in other words the people that perform the job themselves [2]. The “human factor” is one of the most important aspects to be considered when a new project is conceived.

In this way, it is possible to characterise software development as a job with high levels of superposition, i.e., a lot of people perform the same task at the same time or a single person performs several tasks simultaneously. At the same time the developers are multi assigned, performing the jobs of different projects in short time intervals. To this it is necessary to add timetable flexibility, according to interruptions during work, extra hours and unpunctuality, and the fuzzy definition of the aims to be achieved. Therefore, uncertainty management will be a constant feature in project management.

Some methodologies have been defined which help professionals to perform their jobs with a set of good practices of Software Engineering. One of the most prominent techniques is the Personal Software Process (PSP) [3]. Inside the PSP, there are defined methods which enable carrying out better planning and tracing of the work, establishing measurable aims and offering the possibility of using a well-defined and measured working process.

Facing the problem of software project estimation, taking into consideration the human factor, in this paper several soft-computing techniques are proposed, such as fuzzy prototypes and clustering algorithms with the aim of facilitating the modelling of time management behaviour of a software engineer. As a data source to perform this job of knowledge extraction, the techniques and tools defined as PSP are established, always bearing in mind the difficulty that exists when collecting real basic data of work on this methodology. The main purposes of this work are the following:

1) To detect patterns and extract conclusions in order to construct estimations based on knowledge.
2) To provide the required tools to perform adaptive programming on the tasks making it possible to arrange the tasks which a component of the project must perform during a temporary cycle.

The remainder of the work is organised as follows: Section 2 provides an overview of the PSP method and tools, and a brief description of related work. The processing steps of knowledge extraction are described in detail, from the specification of the data sources to the construction of the fuzzy prototypes in section 3. In section 4, the case study from which the knowledge has been extracted is developed. Finally, in section 5, some conclusions and future work are outlined.

II. RELATED WORK

The PSP introduced in [4] provides a structured framework of forms, guidelines, and procedures for software development in order to help software engineers to be more productive and produce higher quality software.
developer behaviour has been achieved according to the kind of working days or weeks, as well as the personal features of the professional. With this aim, techniques as clustering algorithms and Fuzzy Deformable Prototypes have been used.

The next step is the real application of this extracted knowledge in the estimation and scheduling of new projects. Our aim is to minimise stress while ensuring that work proceeds towards maximal productivity and satisfaction. We also aim to extract knowledge about data from other studies about simulators and controlled experiments with the objective of being able to define more concisely the possible behaviour of software developers.

ACKNOWLEDGEMENT

This research is partially supported by the projects TIN2007-67494 of the Spanish Science and Innovation Ministry and PEIC09-0196-3018 of the Regional Government of Castilla-La Mancha.

REFERENCES