Available online at www.sciencedirect.com

ScienceDirect

International Journal of Project Management 33 (2015) 1597 – 1609

International Journal of Project Management

www.elsevier.com/locate/ijproman

Stochastic earned value analysis using Monte Carlo simulation and statistical learning techniques

Fernando Acebes [a,*], María Pereda [b], David Poza [a], Javier Pajares [a], José Manuel Galán [b]

[a] INSISOC, Departamento de Organización de Empresas y Comercialización e Investigación de Mercados, Escuela de Ingenierías Industriales, Universidad de Valladolid, Paseo del Cauce 59, Valladolid 47011, Spain

[b] INSISOC, Área de Organización de Empresas, Departamento de Ingeniería Civil, Escuela Politécnica Superior, Universidad de Burgos, Edificio La Milanera, C/Villadiego S/N, Burgos 09001, Spain

Received 19 February 2015; received in revised form 22 June 2015; accepted 29 June 2015
Available online 21 July 2015

Abstract

The aim of this paper is to describe a new integrated methodology for project control under uncertainty. This proposal is based on Earned Value Methodology and risk analysis and presents several refinements to previous methodologies. More specifically, the approach uses extensive Monte Carlo simulation to obtain information about the expected behavior of the project. This dataset is exploited in several ways using different statistical learning methodologies in a structured fashion. Initially, simulations are used to detect if project deviations are a consequence of the expected variability using Anomaly Detection algorithms. If the project follows this expected variability, probabilities of success in cost and time and expected cost and total duration of the project can be estimated using classification and regression approaches.

Keywords: Project Management; Earned Value Management; Project control; Monte Carlo simulation; Project risk management; Statistical learning; Anomaly Detection

1. Introduction

Project control and monitoring involve comparing a plan or baseline with the actual performance of the project. The analyses of these deviations are aimed at taking actions, if needed, to early correct the possible problems that can put in danger the objectives of the plan. The most popular managerial methodology used in Project Management is Earned Value Management (EVM). This framework integrates in a unified approach, three dimensions of the project – scope, time and cost – using monetary units as common pivotal measure (Abba and Niel, 2010; Anbari, 2003; Blanco, 2013; Burke, 2003; Cioffi, 2006; Fleming and Koppelman, 2005; Henderson, 2003, 2004; Jacob, 2003; Jacob and Kane, 2004; Kim et al., 2003; Lipke, 1999, 2003, 2004; McKim et al., 2000).

Recent research enhances the standard approach to EVM incorporating statistical analysis, learning curves or fuzzy set theory, especially for project predictions at completion (Colin and Vanhoucke, 2014; Hazir, 2015; Lipke et al., 2009; Moslemi Naeni and Salehipour, 2011; Naeni et al., 2011; Narbaev and De Marco, 2014; Plaza and Turetken, 2009; Tseng, 2011; Wauters and Vanhoucke, 2014). An active area of development is currently focused on integration of EVM with risk management analysis. Progress in this line has produced decision tools based on two metrics to estimate if the deviations may be caused by structural problems or if they are compatible with the expected range of variability, derived from the stochastic nature of the project – estimated variability of costs and durations of project activities – (Pajares and López-Paredes, 2011). These results have been refined using Monte Carlo simulation and statistical control charts (Acebes et al., 2014).

In this work, we initially investigate the alternative use of Anomaly Detection algorithms to detect structural deviations

* Corresponding author.

http://dx.doi.org/10.1016/j.ijproman.2015.06.012
0263-7863/00/

in projects. Assuming the stochastic definition of the project, we then advance the research proposing the use of statistical learning techniques and Monte Carlo simulation to estimate the probability of over-runs (delays or over-costs) and the success decision boundaries. The analysis is completed with additional methodologies to predict not only over-runs but also the expected budget and time.

The rest of the paper is organized as follows. First, in the Background section, we review previous methodologies related to the same problem we face in this research. Then in the Methodology section, we explain the statistical learning methodologies we use in our framework, and how we can apply them for project control. In particular, we explain the role of Anomaly Detection algorithms, and classification and regression techniques. Finally, we address a case study to show how these methodologies work together with EVM for project control.

2. Background

EVM does not provide a way to determine whether deviations are due to the expected range of variability associated with the stochastic nature of the project or they may be caused by unexpected events affecting project runtime. Knowing the reasons for project over-runs would provide the project manager with valuable information for decision-making. Concretely, the fact of the deviations exceeding the expected variability would be a warning sign that the project realization is probably not running as planned. Being aware of this situation would allow applying actions to redirect the evolution of the project.

With this idea in mind, previous research has provided two different frameworks based on EVM that inform whether deviations are within the probabilistic expected level or not: the Schedule Control Index (SCoI)/Cost Control Index (CCoI) framework (Pajares and López-Paredes, 2011) and, more recently, the Triad Methodology (Acebes et al., 2014).

2.1. The Schedule Control Index (SCoI) and Cost Control Index (CCoI) framework

Pajares and López-Paredes (2011) use Monte Carlo simulation to obtain the statistical distribution of the cost and the duration at the end of the project. This information is used to select the confidence level (both in terms of time and in terms of cost) that will be used to monitor the performance of the project. If the cost (time) at the end of a particular project is below the cost (time) at the selected confidence level, the cost overrun (delay) is considered to be caused by the randomness of the real costs (durations) of the activities. Therefore, the difference between the project cost (duration) at the confidence level and the mean project cost (duration) gives an idea of the maximum deviation that can be explained by the stochastic nature of the cost (duration) of the activities. In other words, this difference is considered as the size of the cost/time buffers for the project.

Nevertheless, knowing the size of these buffers at the very end of the project is useless since it does not permit to make decisions that rectify the project performance during runtime. It would be desirable, however, to know what portion of these buffers is available at any particular time during the project execution. To this aim, the authors develop a system to redistribute the size of project cost and time buffers throughout the project life cycle. In order to determine the portion of buffer that corresponds to each period, the authors define the project risk baseline as the residual uncertainty to complete the remaining activities of the project. Then, every time interval is provided with a portion of the cost/time buffers namely $ACBf_t$ and $ASBf_t$ respectively. The size of these buffers is proportional to the risk eliminated between two consecutives periods (i.e. the difference between two adjacent points in the risk baseline).

The authors define two control indices based on these buffers: Cost Control Index (CCoI) and Schedule Control Index (SCoI), which are equal to the traditional indices used by EVM: schedule variance (SV) and cost variance (CV) plus the corresponding portion of the cost/time buffer $ACBf_{t=ES}/ASBf_t$:

$$SCoI_t = ASBf_t + SV$$

$$CCoI_t = ACBf_{t=ES} + CV.$$

Therefore, the new criteria to diagnose the time performance of the project are the following: the project is behind schedule if $SV < 0$ (as in the traditional EVM). However, depending on the value of $ASBf_t$, this delay may be due to by the randomness of the real duration of the activities or caused by structural problems. If $ASBf_t$ is greater than SV (and thus $SCoI_t > 0$), we infer that the delay falls within the expected variability. However, if SV is greater than $ASBf_t$ (and thus $SCoI_t < 0$), the delay may be caused by structural problems and thus require measures to redirect the performance of the project.

Similarly, the index $CCoI_t$ warns about cost overruns ($CV > 0$) and, when they occur, it reports whether the overruns are within the expected variability ($CCoI_t > 0$) or not ($CCoI_t < 0$).

2.2. The Triad Methodology: (x, t, c)

Acebes et al. (2014) developed a different method to determine whether the project deviations are within the expected variability or whether, on the contrary, they are due to undiscovered factors affecting the project performance. This method also uses Monte Carlo simulation to obtain the statistic distribution of all the possible realizations of a project. However, unlike the method shown above, the authors directly determine the statistical distribution of cost and time at intermediate percentages of completion of the project.

For every realization of a Monte Carlo simulation, the system provides a final cost and time. That is, when the percentage of completion of a simulation is 100% (x = 100%), we obtain the final cost for that simulation ($c_{100\%}$) and the final duration for that simulation ($t_{100\%}$). This triad (100%, $t_{100\%}$, $c_{100\%}$) gives a name to this methodology. Afterwards, the algorithm calculates the cost and time at the desired intermediate time intervals for that particular simulation. It is important to mention that the percentage of completion of the project is calculated in terms of cost. This means that, for example, the project is 50% completed at the time when