Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 55 (2016) 41 – 46

www.elsevier.com/locate/procedia

5th CIRP Global Web Conference Research and Innovation for Future Production

An approach for selecting cost estimation techniques for innovative high value manufacturing products

Oliver Schwabe[a*], Essam Shehab[a] and John Erkoyuncu[a]

[a]*Cranfield University, Building 50, Cranfield, Bedford MK43 0AL, U.K.*

* Corresponding author. Tel.: +49 (0) 170 9053671;. *E-mail address:* o.schwabe@cranfield.ac.uk

Abstract

This paper presents an approach for determining the most appropriate technique for cost estimation of innovative high value manufacturing products depending on the amount of prior data available. Case study data from the United States Scheduled Annual Summary Reports for the Joint Strike Fighter (1997-2010) is used to exemplify how, depending on the attributes of a priori data certain techniques for cost estimation are more suitable than others. The data attribute focused on is the computational complexity involved in identifying whether or not there are patterns suited for propagation. Computational complexity is calculated based upon established mathematical principles for pattern recognition which argue that at least 42 data sets are required for the application of standard regression analysis techniques. The paper proposes that below this threshold a generic dependency model and starting conditions should be used and iteratively adapted to the context. In the special case of having less than four datasets available it is suggested that no contemporary cost estimating techniques other than analogy or expert opinion are currently applicable and alternate techniques must be explored if more quantitative results are desired. By applying the mathematical principles of complexity groups the paper argues that when less than four consecutive datasets are available the principles of topological data analysis should be applied. The preconditions being that the cost variance of at least three cost variance types for one to three time discrete continuous intervals is available so that it can be quantified based upon its geometrical attributes, visualised as an n-dimensional point cloud and then evaluated based upon the symmetrical properties of the evolving shape. Further work is suggested to validate the provided decision-trees in cost estimation practice.

Keywords: Computational complexity; Cost estimate uncertainty; Method selection

1. Introduction

This paper presents an approach for determining the most appropriate technique for cost estimation of innovative high value manufacturing products depending on the amount of a priori data available [1]. High value manufacturing products are understood as such products which are the results of "…the application of leading edge technical knowledge and expertise…" and result in "…the creation of products, production processes, and associated services which have strong potential to bring sustainable growth and high economic value…" [2]. Exemplary high value manufacturing products are aerospace engines and airframes, sea vessels and defence ground vehicles. Innovation is declared to exist when no verified and accurate cost models are available. Under the associated conditions of deep uncertainty and small data it is especially the plethora of varying plausible future scenarios that calls for a deeper understanding of available methods. Uncertainty is defined as unintended cost variance with a probability of 100% and an unknown impact [3]. The presented approach provides guidelines for when the following cost estimation techniques can be applied with confidence: analogy or expert opinion, topological estimation, parametric estimation and standard statistical regression. The cost estimation technique based upon topological estimation is novel in the field of cost estimation and deemed suitable for

doi:10.1016/j.procir.2016.08.017

filling the gap between zero and four discrete time elements of prior data [4]. The guidelines are based upon the complexity of a priori data as defined by Kolmogorov [5] and applied to an evolving bit string describing the propagation direction of cost variance metrics deemed relevant by the estimator (i.e. the direction of cost variance propagation).

Section 2 describes how the principles of Kolmogorov complexity are used to determine the time windows for differing forecast techniques and Section 3 presents the case study data. Section 4 presents the fundamental selection guidelines and Section 5 presents the process for determining which forecast technique is most applicable when. Section 6 discusses the concept of innovativeness and Section 7 shares a series of common estimation situations and describes how the presented approach leads to specific forecast technique recommendations. Section 8 shares a conclusion and provides recommendations for future research.

2. Complexity for determining time windows

It is suggested that Kolmogorov complexity [5] is a suitable indicator for determining when a specific forecasting technique is applicable or not. The metric of Kolmogorov complexity signifies the degree of compression a binary string can be subject to whereby compression is understood as the process of converting the sequence of bits into the description of the pattern represented by that bit sequence. The bit sequence is hence transformed into a program that can generate exactly that bit sequence. The program consists of a descriptor language which explains how a sequence of instructions is applied by a Turing Machine in order to generate the bit string.

The data of interest is the prior data related to the financial cost variance of the high value manufacturing product, specifically across at least three dimensions of cost variance such as cost changes due to variance in requirements, schedule or units ordered. This data needs to cover iterative discrete time intervals prior to the point in time where the cost estimate is being performed.

The first boundary suggested by Kolmogorov complexity is that the data from at least 42 discrete time intervals is required before pattern recognition approaches can be applied for forecasting purposes. This includes the application of standard regression techniques [5].

The second boundary suggested by Kolmogorov complexity is that depending on the length of the bit string the actual complexity score of individual bit strings can be grouped into groups of identical complexity. Bit strings of length one or two have the same complexity group. It is first the bit strings with a length of three elements that demonstrate this behavior. It is then with the fourth element that a first determination of stability can be made. The authors therefore suggest that while at time interval zero no techniques other than analogy or expert opinion are feasible, starting with the fourth time intervals parametric models (that depend on an understanding of cost estimating relationships) are applicable. From one to three elements a gap exists that the authors propose to fill with the technique of topological estimation.

3. Case study data

The data selected for exemplifying the selection technique is drawn from the United States Scheduled Annual Summary Reports [6] for the Joint Strike Fighter in the time period of 1997 to 2010 as illustrated by Table 1.

Table 1. Case study data

Baseline	Year	Quantity	Schedule	Engineering	Estimating	Other	Support
1994	1997	0	0	0	140	0	0
1994	1998	0	0	1121	105	0	0
1994	1999	0	0	1121	105	0	0
1994	2000	0	0	1121	105	0	0
2002	2001	0	19	5452	7438	0	0
2002	2002	16249	0	2458	2330	0	2595
2002	2003	16249	8024	4370	17153	0	2735
2002	2004	16249	8139	7940	11998	0	5092
2002	2005	16249	8208	8279	17838	0	8054
2002	2006	16249	8797	9687	18849	0	11218
2002	2007	16249	8797	9687	30738	0	58
2002	2008	16249	8797	9687	30738	0	58
2002	2009	16119	8797	9687	52380	0	6753
2002	2010	16119	8797	9687	77984	0	13151

The “baseline” is the year in which the technical baseline estimate was created. The “year” is the discrete time period for which accounting data was available. The factors “quantity” through “support” are the reasons for cost variance assessed in the accounting period and the numbers entered represent the total absolute financial variance in USD$ million as compared to the baseline estimate. This variance may represent cost increases or reductions whereby the focus of the research is absolute variance from target.

As illustrated in Table 2 the case study data is now analysed to determine whether the cost variance for an accounting time period is higher (“1”), lower (“0”) or equal (“0”) to the previous time period. If the year of the baseline estimate changes a “1” is also assigned.

Table 2. Case study data

Year	Quantity	Schedule	Engineering	Estimating	Other	Support	Total
1997	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1998	0	0	0	0	0	0	1
1999	0	0	0	0	0	0	0
2000	0	0	0	0	0	0	0
2001	1	1	1	1	1	1	1
2002	1	0	0	0	0	1	1
2003	0	1	1	1	0	1	1
2004	0	1	1	0	0	1	1
2005	0	1	1	1	0	1	1
2006	0	1	0	1	0	1	1
2007	0	0	0	1	0	0	1
2008	0	0	0	0	0	0	0
2009	0	0	0	1	0	1	1
2010	0	0	0	1	0	1	1