Computers & Industrial Engineering 112 (2017) 654–662

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

A bootstrap method for uncertainty estimation in quality correlation algorithm for risk based tolerance synthesis

Raed S. Batbooti [a], R.S. Ransing [a,*], M.R. Ransing [b]

[a] College of Engineering, Swansea University, Swansea SA1 8EN, United Kingdom
[b] p-matrix Ltd, Swansea, United Kingdom

ARTICLE INFO

Article history:
Available online 3 September 2016

Keywords:
7Epsilon
Six Sigma
No-Fault-Found product failures
Bootstrapping
In-tolerance faults and in-process quality improvement

ABSTRACT

A risk based tolerance synthesis approach is based on ISO9001:2015 quality standard's risk based thinking. It analyses in-process data to discover correlations among regions of input data scatter and desired or undesired process outputs. Recently, Ransing, Batbooti, Giannetti, and Ransing (2016) proposed a quality correlation algorithm (QCA) for risk based tolerance synthesis. The quality correlation algorithm is based on the principal component analysis (PCA) and a co-linearity index concept (Ransing, Giannetti, Ransing, & James, 2013). The uncertainty in QCA results on mixed data sets is quantified and analysed in this paper.

The uncertainty is quantified using a bootstrap sampling method with bias-corrected and accelerated confidence intervals. The co-linearity indices use the length and cosine angles of loading vectors in a p-dimensional space. The uncertainty for all p-loading vectors is shown in a single co-linearity index plot and is used to quantify the uncertainty in predicting optimal tolerance limits. The effects of re-sampling distributions are analysed. The QCA tolerance limits are revised after estimating the uncertainty in limits via bootstrap sampling. The proposed approach has been demonstrated by analysing in-process data from a previously published case study.

1. Introduction

1.1. Risk based tolerance synthesis

The clause 6.1 of ISO9001:2015 quality standard requires organisations to continually improve the process by enhancing the occurrence of desired process outputs and reducing or preventing the instances when the process has produced undesired results. Except for robust processes, the variation in process inputs may lead to deviation from expected or desired results. The relationship between process inputs and outputs is normally too complex for the tolerance synthesis problem to be modelled by the underlying physics alone. Firstly, the governing equations used to model the physics may not describe the real model accurately and secondly, we may not even know the underlying physics sufficiently (Lewis & Ransing, 2000; Lewis, Ransing, Pao, Kulasegaram, & Bonet, 2004; Pao, Ransing, Lewis, & Lin, 2004; Postek, Lewis, Gethin, & Ransing, 2005). In a continuously monitored manufacturing environment, the synthesis of in-process data can help process engineers to discover subtle relationships among process inputs and outputs (Lewis & Ransing, 1997). Tolerance synthesis is the process of determining allowable variation in products and processes in order to meet the quality requirements (Li, Kokkolaras, Papalambros, & Hu, 2008). For a multi-process manufacturing system, it is essential that the variability in all process inputs (including interactions among process inputs) is analysed to study the variation in one or more process outputs. The tolerance synthesis is usually based on a mathematical model that describes the variation of the process inputs (Ding, Jin, Ceglarek, & Shi, 2005). A penalty matrix approach is used to estimate the deviation from expected results (Ransing, Giannetti, Ransing, & James, 2013). To embed the risk based thinking in a tolerance synthesis problem, process responses are categorised into three categories (i) desired, (ii) unacceptable and, (iii) a middle region between the two categories. A zero penalty value is assigned to the desired response and a 100 penalty value is given for the unacceptable process response. A process response in the middle region is assigned a penalty value between zero and hundred. A correlation between factor values and penalty values for a given response is discovered using a principal component analysis (PCA) based co-linearity index (CLI) plot (Ransing et al., 2013). The length and angle of each loading vector is calculated in a reduced p-dimensional subspace and is used in a CLI plot. The quality correlation algorithm (QCA)

* Corresponding author.
E-mail address: r.s.ransing@swansea.ac.uk (R.S. Ransing).

http://dx.doi.org/10.1016/j.cie.2016.09.002
0360-8352/

(Ransing, Batbooti, Giannetti, & Ransing, 2016) discovers optimal tolerance limits by projecting scores on correlated factors and responses. The scores, bounded between a factor direction and the corresponding response direction in a CLI plot, are used to calculate new tolerance limits for the corresponding factor. For quantitative variables, the range of factor values corresponding to chosen scores defines a new tolerance limit. On the other hand, the optimal and avoid categories for categorical variables are determined by calculating the percentage of occurrences for a corresponding category in the collected scores vector. The obtained projected scores vector leads to an optimal percentage of occurrences if the variable is correlated with low penalty values. For an avoid range, the variable correlates with high penalty values. It is suggested that these new tolerance limits be included in a modified process failure modes effect analysis (PFMEA) table in order to create a reusable organisational knowledgebase (Batbooti, Ransing, & Ransing, 2015).

The number of in-process observations, available for undertaking a tolerance synthesis project, is normally very small (~50–100). The number of input and output variables are of the similar size of number of observations (~50–100) (Ransing et al., 2013). The small sample size can affect the reliability of model predictions, and hence a measure needs to be developed to quantify uncertainty in the model. In the tolerance synthesis context, bootstrapping is based on the notion that the in-process data is representative of the entire population of the data set as the sample size increases to infinity. The QCA estimates population parameters such as the upper and lower tolerance limits for each factors. The novelty and originality of this work is in extending the algorithm proposed by Timmerman, Kiers, and Smilde (2007) for quantifying the uncertainty in the QCA. The bootstrap parameters used by Timmerman et al. (2007) are different to those used in the QCA. Resampling from the in-process data set is used to imitate the sampling process from the population. The bootstrap parameter distribution is used to quantify the uncertainty by calculating standard errors and confidence intervals. The revised upper and lower tolerance limits of the QCA are derived from the bootstrap parameter values using the weighted mean formulation proposed by Grela (2013) and Finch (2009).

1.2. Uncertainty estimation with a bootstrap resampling

Timmerman et al. (2007) have compared the estimation of confidence intervals using the bootstrap approach as well as the asymptotic approach. It was shown that the bootstrap approach was better suited for predicting uncertainty and hence the confidence intervals. The methodology for calculating standard errors and confidence intervals in bootstrap procedures is discussed widely in the literature (Efron, 1977; Efron & Tibshirani, 1993; Hastie, Tibshirani, & Friedman, 2009; Wehrens, Putter, & Buydens, 2000).

A PCA bootstrap method has been used to find the variances of PCA loadings (Chatterjee, 1984; Lambert, Wildt, & Durand, 1990, 1991). A thousand bootstrapped samples were generated with replacement to determine stopping criteria for choosing number of PC's (Jackson, 1993). Further studies on the finding of the number of retained PCs have been discussed in the literature (Besse, 1992; Daudin, Duby, & Trecourt, 1988; Peres-Neto, Jackson, & Somers, 2005).

Smith and Gemperline (2002) compared two parametric bootstrap methods for analysing small data sets in order to improve the estimation of misclassification rates of microcrystalline cellulose.

A non-parametric bootstrap method was used in an exploratory factor analysis to estimate results of Procrustes rotation to a target pattern matrix (Raykov & Little, 1999). Bootstrap confidence intervals were estimated for scores and loading values, as well as the global clusters in PCA, to assess the uncertainty (Babamoradi, van den Berg, & Rinnan, 2013). The study was conducted on two small datasets.

A bootstrap based method is proposed to enhance QCA results by estimating uncertainty in the algorithm for solving risk based tolerance synthesis problems. Table 1 illustrates the symbols used in the paper. The proposed uncertainty estimation method is described in Section 2. A study of one thousand bootstrapped samples is discussed in Section 3 and Section 4 concludes the paper.

2. Bootstrap uncertainty estimation based on the QCA

2.1. The quality correlation algorithm (QCA)

In a given timeframe, each occurrence of process result is recorded and assessed as desired or undesired process outcome. The deviation from the expected results quantified with a penalty value. For a continuous monitoring environment, during the same timeframe the factor values are normally measured at a much higher frequency rate. The median, the average of top the 5% of values and the average of the bottom 5% of values is determined for each factor using values collected in the timeframe and uniquely associated with the penalty value for the given timeframe. In the 7Epsilon context, this dataset is referred to as an equal frequency rate data set. This dataset is stored in matrix X and is referred to as the in-process data matrix with m number of observations and n process variables. The process variables include categorical and quantitative factors and one or more process responses. The data pre-treatment proposed by Giannetti et al. (2014), and shown in Table 2, is applied to the in-process data matrix X to transform this matrix to XT. As shown in Fig. 1, this is the first step of the QCA. The second step applies the PCA to the transformed in-process data

Table 1
Nomenclature.

$\hat{se}$	Boot strap standard error
$\hat{a}$	The acceleration or skewness constant
$\hat{z}_0$	The bias correction
θ	The static of interest
B	Number of bootstraps
D_e	Diagonal matrix containing the square roots of eigenvalues
D_s	Diagonal matrix containing the standard deviations of the columns of XT
E	Error matrix
HB	Higher the better penalty value settings
L	Loading matrix
LB	Lower the better penalty value settings
LL^j	Lower (minimum) value of x_o^j
n_c	Number of correlated parameters resulted from applying CLI
n_x^j	Length of vector x_o^j
Q	Is the number of original categorical variables
R	Response
s_j	Standard deviation of factor j
$t^{\#,j}$	The scores vector containing scored bounded between the loading vector j direction and the response direction
$T^{\#}$	Projected score matrix
T	Score Matrix
TL^j	Tolerance limit for a factor j
UL^j	Upper (maximum) value of x_o^j
V	Matrix of eigenvectors containing eigenvectors as column vectors ordered by greatest eigenvalues
X	Original data matrix with ($m \times n$) dimensions, with m is the number of observations and n number of variables
x_o^j	Vector with (o: either optimal or avoid) depending on direction of variable
X_{Recon}	Reconstruct matrix X from PCA model parameters
XT	Data matrix after pre-treatment
z_j	Dummy variable taking binary value one if the categorical variable has been observed and zero otherwise