European Journal of Operational Research 237 (2014) 857–870

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Production, Manufacturing and Logistics

A hybrid wrapper–filter approach to detect the source(s) of out-of-control signals in multivariate manufacturing process

Shamsul Huda [a,*], Mali Abdollahian [b], Musa Mammadov [a], John Yearwood [c], Shafiq Ahmed [b], Ibrahim Sultan [d]

[a] Center for Informatics and Applied Optimization, School of SITE, University of Ballarat, VIC 3353, Australia
[b] School of Mathematics and Geospatial Science, RMIT University, VIC, Australia
[c] Center for Informatics and Applied Optimization (CIAO), School of SITE, University of Ballarat, VIC 3353, Australia
[d] School of SITE, University of Ballarat, VIC 3353, Australia

ARTICLE INFO

Article history:
Received 6 November 2012
Accepted 13 February 2014
Available online 4 March 2014

Keywords:
Multivariate control chart
Fault diagnosis
Global optimization
Wrapper and filter approaches

ABSTRACT

With modern data-acquisition equipment and on-line computers used during production, it is now common to monitor several correlated quality characteristics simultaneously in multivariate processes. Multivariate control charts (MCC) are important tools for monitoring multivariate processes. One difficulty encountered with multivariate control charts is the identification of the variable or group of variables that cause an out-of-control signal. Expert knowledge either in combination with wrapper-based supervised classifier or a pre-filter with wrapper are the standard approaches to detect the sources of out-of-control signal. However gathering expert knowledge in source identification is costly and may introduce human error. Individual univariate control charts (UCC) and decomposition of T^2 statistics are also used in many cases simultaneously to identify the sources, but these either ignore the correlations between the sources or may take more time with the increase of dimensions. The aim of this paper is to develop a source identification approach that does not need any expert-knowledge and can detect out-of-control signal in less computational complexity. We propose, a hybrid wrapper–filter based source identification approach that hybridizes a Mutual Information (MI) based Maximum Relevance (MR) filter ranking heuristic with an Artificial Neural Network (ANN) based wrapper. The Artificial Neural Network Input Gain Measurement Approximation (ANNIGMA) has been combined with MR (MR-ANNIGMA) to utilize the knowledge about the intrinsic pattern of the quality characteristics computed by the filter for directing the wrapper search process. To compute optimal ANNIGMA score, we also propose a Global MR-ANNIGMA using non-functional relationship between variables which is independent of the derivative of the objective function and has a potential to overcome the local optimization problem of ANN training. The novelty of the proposed approaches is that they combine the advantages of both filter and wrapper approaches and do not require any expert knowledge about the sources of the out-of-control signals. Heuristic score based subset generation process also reduces the search space into polynomial growth which in turns reduces computational time. The proposed approaches were tested by exhaustive experiments using both simulated and real manufacturing data and compared to existing methods including independent filter, wrapper and Multivariate EWMA (MEWMA) methods. The results indicate that the proposed approaches can identify the sources of out-of-control signals more accurately than existing approaches.

1. Introduction

Multivariate process control techniques were established by Hotelling in his paper Hotelling (1947). He introduced the problem of correlation between the quality characteristics of a process and devised the well-known T^2 statistic to identify whether the whole process is out of control. Hotelling's T^2 statistic is the optimal test statistic for detecting a general shift in the process mean vector for an individual multivariate observation. However, the technique has several practical drawbacks. Critically, when the T^2 statistic

* Corresponding author. Tel.: +61 3 53276217.
E-mail addresses: s.huda@ballarat.edu.au (S. Huda), mali.abdollahian@rmit.edu.au (M. Abdollahian), m.mammadov@ballarat.edu.au (M. Mammadov), j.yearwood@ballarat.edu.au (J. Yearwood), shafiq.ahmad@rmit.edu.au (S. Ahmed), i.sultan@ballarat.edu.au (I. Sultan).

http://dx.doi.org/10.1016/j.ejor.2014.02.032
0377-2217/

indicates that a process is out of control, it does not provide information on which variable or set of variables is out of control. Moreover, it is difficult to distinguish location shifts from scale shifts since the T^2 statistic is sensitive to both types of process changes.

The difficulty of interpreting an out-of-control signal on a multivariate control chart has been discussed extensively Alt (1985), Doganaksoy, Faltin, and Tucker (1991), Murphy (1987), Pignatiello and Runger (1990), Lowry, Woodall, Champ, and Rigdon (1992) and Linderman, McKone-Sweet, and Anderson (2005) among others. When two or more correlated variables are monitored, use of a multivariate chart may cause signals at opposing times to the signals given by a set of univariate charts on the individual variables. This is because the control region for a multivariate chart on correlated variables is represented by a tilted elliptical region as opposed to the non-tilted square region obtained by the use of separate charts. In fact, the use of separate charts does not allow for the information concerning the correlation of the variables to be utilized. However, the combination of using a multivariate control chart for signalling purposes and then using separate charts for diagnostic purposes is often effective.

Given an m-dimensional quality characteristics set of data, a fault detection and diagnosis subsystem in a multivariate process needs to find the optimal sources of fault for the out-of-control signals from the 2^m subsets of quality characteristics which is computationally expensive and exponential to the dimensions of characteristics set (Jackson & Morris, 1957). Moreover, the performance of the algorithms depends on its evaluation criterion and search strategies.

Murphy (1987) proposed a method to identify the out-of-control variables based on discriminant analysis. He divided the complete set of variables into two subsets and then tried to determine which one caused an out-of-control signal. Alt (1985) proposed the use of the univariate t-statistic for ranking the variables most likely to have changed. Then, to further strengthen the belief that a certain variable has changed, they applied the Bonferroni-type interval. The obvious drawback of this method is that it only tells you which variable is most likely to have shifted, which is not conclusive. Also, this method does not allow the user to study the trends.

Mason, Tracy, and Young (1995) showed that signal interpretation of the T^2 statistic is greatly aided if the corresponding value is partitioned into independent parts. The characteristic that is significantly contributing to the signal is more readily identified by decomposing the T^2 statistic into independent parts, each of which reflects the contribution of an individual variable. The drawback of this method is the extensive computation and its sensitivity to the number of variables. This approach reduces the search space from exponential to $O(m^2)$. However, for a large number of quality characteristics, search space still would be a cumulative factor for computational time.

Many approaches have been suggested for identifying the variable or group of variables that causes the out-of-control signals. They can be grouped broadly into three main categories: (1) the wrapper model with expert knowledge (Chen & Wang, 2004; Chen & Wang, 2004; Francisco & Jos, 2010; Jianbo, Lifeng, & Xiaojun, 2009; Low, Hsu, & Yu, 2003; Zorriassatinea, Tannockb, & O'Brienb, 2003) (2) filter models with wrapper evaluation (Rodger, 2012; Sylvain, Teodor, & Abdessamad, 2008; Wang & Dub, 2000), and (3) statistical approaches (Francisco & Jos, 2010; Mason et al., 1995).

Wrapper approach is one of the main approaches for fault diagnosis that uses a pre-determined induction algorithm where the performances of the algorithms are used as the evaluation criteria, see for example (Chen & Wang, 2004; Chen & Wang, 2004; Deborah, Christopherand, & Wolfe, 2006; Francisco & Jos, 2010; Jianbo et al., 2009; Lorton, douladirad, & Grail, 2013; Low et al., 2003; Wang, 2012; Zorriassatinea et al., 2003). The wrapper models use expert knowledge to build training pattern which is capable of identifying the sources of future out-of-control signals. The main disadvantage of these approaches is that they need expert inputs about the sources of out-of-control signals which may introduce human error. In addition, the wrapper models face huge computational overhead due to the use of the induction algorithm's performance and subset generation process used in the wrapper search space. None of the above approaches use any heuristic score in subset generation process to reduce the search space for fault diagnosis in the multivariate processes.

Outside the multivariate process control domain, wrapper approaches have also been used in many classification problems. Let us refer to few recent articles by Kohavi and John (1997), Puronnen, Tsymbal, and Skrypnik (2000), Huang, Z Cai, and Xu (2008), Hsu, Huang, and Schuschel (1999), Romero and Sopena (2008). In particular, Hsu et al. (1999) proposed a wrapper based heuristic in artificial neural network (ANN) known as Artificial Neural Network Input Gain Measurement Approximation (ANNIGMA) for a general classification problem. Hsu et al. (1999) demonstrated that a wrapper-based heuristic can improve search performance in a standard training environment of an ANN. However, the approach (Hsu et al., 1999) also needs expert-knowledge for ANN training and ignores the problem of standard training in ANN. The other approaches use different search strategies for subset generation based-on wrapper performance function which involves numerous call of wrapper re-training in a supervised experimental environment.

Filter models are one of the computationally cheap models which have broadly been considered for many fault diagnosis problems, for example by Sylvain et al. (2008), Wang and Dub (2000), Jackson and Morris (1957). In contrast to wrapper models, filter models involve the implementation of a search algorithm by using different heuristics to the data set prior to its use for induction algorithm. Diverse heuristics have been reported in the literature, for example, Principal Component Analysis (PCA) by Jackson and Morris (1957) and Mutual Information by Wang and Dub (2000). Heuristics of filter models estimate the discrimination capabilities of the subsets of the quality characteristics which is followed by the ranking of the source characteristics based on the different search strategies and wrapper classifier. Filter approaches have also been applied in many other problems, for example, bankruptcy prediction, medical diagnosis Gene Selection, malware detection. Let us refer to few recent articles by Santos et al. (2010), Santos, Devesa, Brezo, Nieves, and Bringas (2012), Guo and Lyu (2006), Tsai (2009), Krier, Francois, Wertz, and Verleysen (2006), Ng and Chan (2005). In particular, Tsai (2009) provide a review of filter approaches and shows their comparative performances. Because filter models are independent from the induction algorithm, the selected quality characteristics subsets may result in poor fault diagnosis prediction accuracies. Principal components (Jackson & Morris, 1957) are not easily interpretable in many cases and do not have a one-to-one relation with the original variables. Nevertheless, in some cases, depending on the context, they can be very useful.

Therefore there is a need for development of a procedure that can avoid the drawbacks of both filter and wrapper fault diagnosis approaches and can reduce the search space for better computational complexity for the detection of the source(s) of out-of-control signals. In this paper, we propose a hybrid wrapper–filter approach by injecting the filter's ranking score in the wrapper approach. The objective is to find a suitable heuristics score for the wrapper stage that can improve the search space complexity and can find most significant compact set of the sources of out-of-control signals without any additional information other than control list for a multivariate quality control environment. We introduce a new heuristic score by hybridizing Mutual Information (MI) based Maximum Relevance (MR) filter ranking heuristics with the Artifi-