Measurement 55 (2014) 15–24

Contents lists available at ScienceDirect

Measurement

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

Fault diagnosis on material handling system using feature selection and data mining techniques

M. Demetgul [a,*], K. Yildiz [b], S. Taskin [c], I.N. Tansel [d], O. Yazicioglu [e]

[a] Department of Mechatronics Engineering, Technology Faculty, Marmara University, Turkey
[b] Department of Computer and Control Education, Technical Education Faculty, Marmara University, Turkey
[c] Department of Electrical and Electronics Engineering, Engineering Faculty, Celal Bayar University, Turkey
[d] Department of Materials and Mechanical Engineering, Engineering Faculty, Florida International University, FL, USA
[e] Department of Industrial Engineering, Design and Engineering Faculty, Istanbul Commerce University, Turkey

ARTICLE INFO

Article history:
Received 4 February 2014
Received in revised form 21 April 2014
Accepted 23 April 2014
Available online 9 May 2014

Keywords:
Servo-pneumatic
Material handling system
Fault diagnosis
Feature selection
Data mining
Dimension reduction
Gustafson–Kessel
k-Medoids

ABSTRACT

The material handling systems are one of the key components of the most modern manufacturing systems. The sensory signals of material handling systems are nonlinear and have unique characteristics. It is very difficult to encode and classify these signals by using multipurpose methods. In this study, performances of multiple generic methods were studied for the diagnostic of the pneumatic systems of the material handling systems. Diffusion Map (DM), Local Linear Embedding (LLE) and AutoEncoder (AE) algorithms were used for future extraction. Encoded signals were classified by using the Gustafson–Kessel (GK) and k-medoids algorithms. The accuracy of the estimations was better than 90% when the LLE was used with GK algorithm.

1. Introduction

The modern manufacturing facilities have to detect the problems, identify their sources and fix them very quickly with very limited man power. Researchers have started development of computational diagnostic tools for the industrial applications in early 1970s by considering this need. Although, various diagnostic tools have been developed by research community and successfully used in industrial applications in last two decades [1,2] still their capabilities are limited. In this study, feasibility of a multipurpose fault detection approach was investigated. The proposed approach used the combinations of the generic dimension reduction methods for feature extraction and classified the encoded data with clustering algorithms.

One of the key components of the automated manufacturing is material handling systems. Pneumatic and hydraulic systems are widely used for material handling. These systems may have hundreds of actuators and sensors. Identification of faulty components and their locations in a very short time is very difficult. Several studies were performed for development of fault diagnostic tools for these systems in the last decade [3]. The studies mainly aimed evaluation of the condition of the cylinders [4] and digitally controlled valves [5]. Other studies focused on detection of leakage of the seals [6–9], friction increase [4,10] and malfunctions [11–14].

Most of the fault diagnostic tools have two components: feature extractor (encoder) and classifier. Some researchers have used the intelligent data analysis

* Corresponding author. Tel.: +90 216 336 57 70.
E-mail address: mdemetgul@marmara.edu.tr (M. Demetgul).

http://dx.doi.org/10.1016/j.measurement.2014.04.037
0263-2241/

techniques for fault diagnostic [15–17]. Support vector machines [18], self-organizing feature maps (SOM) [19], expert systems, neural networks, rough sets and fuzzy logic have been used for classification of data. The computing complexity of feature extraction and learning process have been the main disadvantages of these approaches.

The data of the material handling systems for the fault diagnosis comes from multiple sensors. The data is high-dimensional and nonlinear. While the large number of data from different sensors provide more information, at the same time feature extraction and classification becomes more complex. The dimension reduction methods compress the data automatically, reduce the noise, may extract features for fault diagnostic and minimize required storage.

Clustering algorithms have been used for classification. Fuzzy c-means (FCM) [20] and its variants Gustafson–Kessel (GK) [21] algorithm are popular pattern classification methods. They have been used for fault detection and isolation [22–24], k-medoids [25] is a partitional clustering algorithm and may be used for classification purposes. DM method [26–29] used diffusion semigroups for learning the global characteristics of the data-set. The complex structures were represented at different scales by the help of these semi groups. The eigenfunctions of Markov matrices were effectively used with this purpose. LLE [30] method used an unsupervised learning algorithm to change the dimensions of the data. The algorithm found the neighbors in X space, calculated weights for reconstruction and calculated the embedding coordinates in Y space by using the calculated weights. AutoEncoder (AE) [31] methods use an artificial neural network (ANN) to learn the compact representation of data set. The dimensionality of the data set is reduced by using this ANN. Various multilayer architectures [32,33] and optimization methods [31,34] have been proposed to improve the performance of the ANN.

GK and k-medoids algorithms were used to create the desired number of clusters to partition or classify the data after it was compacted. GK algorithm [21] calculates the center and covariance matrix to represent the clusters [35,36]. They are used during the optimization process. This approach allows identification of ellipsoidal clusters and improves the performance of the method relative to other approaches. k-medoids [37] is another clustering algorithm. The algorithm divides the data into the groups, chooses the data points as medoids. In this study, Clustering and Data Analysis Toolbox [38] was used for classification of the compressed data.

Diffusion map, AutoEncoder, and Local Lineer Embedding techniques were used for dimension reduction process respectively. In classification process two algorithms were used namely; k-medoids and GK. These algorithms were given in detailed below.

2. Classification and feature extraction

2.1. k-Medoids

It is a standard clustering algorithm [37] where the update rule always moves the cluster center to the nearest data point in the cluster. k-Medoids is a partitioning technique of clustering that clusters the data set of n objects into k clusters with k known a priori. t could be more robust to noise and outliers as compared to k-means because it minimizes a sum of general pairwise dissimilarities instead of a sum of squared Euclidean distances.

2.2. Gustafson–Kessel algorithm (GK)

Gustafson–Kessel algorithm (GK), providing a degree of membership of each data point to a particular cluster creates a fuzzy partition [21]. One set of data to detect clusters of different geometrical shapes, this method is an adaptive distance norm for each cluster was introduced. Each cluster has its own standard distance norm affects the Ai matrix Mi, inducing which statement is the following equation [36].

$$d(x^j, L^i)_M = IIx^j - L^I II = (x^j - L^i)^T M_i (x^j - L^i) \quad (1)$$

possible use of the Z matrix of each cluster, every step of the data fit the geometric structure of the distance norm. Based on the norm-inducing matrices, the aim of the GK method, is obtained by minimizing the function J as Eq. (2).

$$J(P, L, M) = \sum_{i=1}^{c}\sum_{j=1}^{n}(A_i(x^j))^2 d^2(x^j, L^i)'_M \quad (2)$$

$M = (M_1, M, \cdots, M_c)$ is a c-tuple, where the norm-inducing matrices.

2.3. Diffusion maps

Diffusion maps as a system of eigenfunctions of Markov matrices consider effective representation of data geometric descriptions of the original data set to obtain the coordinates [27–29]. A given data set $X = (x, i = 1, \cdots, N)$ is a d-dimensional data space, said N nodes can be built over X is a finite graph corresponding to the N data. Usually, the form of a Gaussian kernel as in Eq. (3)

$$w(x_i, x_j) = \exp(-\frac{IIx_i - x_j II^2}{2a^2}) \quad (3)$$

where σ is the kernel width parameter, is used to construct the similarity matrix. The kernel reflects the degree of similarity between x_i and x_j, and II, II is the Euclidean norm in Rd [39].

2.4. AutoEncoder

Multilayer encoders are hidden layer feed forward neural networks with an odd number [32,33]. The input and the output layer have D nodes and the middle hidden layer has d nodes.

AutoEncoders usually high number of multi-layer connection. Therefore, the back propagation approaches is likely to get stuck in slow convergence and local minimum. In [32] this disadvantage is overcome by performing a Restricted Boltzmann Machines using pretraining (RBMs) [34]. The mean square error between the input and output of the network is trained to minimize (Ideally, the input and output equal). Linear activation functions in the use of neural network, PCA is a very similar AutoEncoder [40].