Expert Systems With Applications 107 (2018) 165–172

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Multi-sensors data fusion through fuzzy clustering and predictive tools

Sagardip Majumder, Dilip Kumar Pratihar*

Department of Mechanical Engineering, Indian Institute of Technology Kharagpur, Kharagpur-721302, India

ARTICLE INFO

Article history:
Received 19 August 2017
Revised 18 January 2018
Accepted 23 April 2018
Available online 24 April 2018

Keywords:
Multi-sensors data
Input-output relationships
Fuzzy clustering
Fuzzy reasoning tool

ABSTRACT

Sensory data are generally associated with imprecision and uncertainty, and consequently, it becomes difficult to extract useful information from them. The problem becomes even more difficult to handle, when the data are collected using multiple sensors. Realizing the ability of fuzzy sets to deal with imprecision and uncertainty, a multi-sensors data fusion technique was developed in this study by using fuzzy clustering and predictive tools. The data were first clustered based on their similarity using an entropy-based fuzzy C-means clustering technique and the obtained clusters were utilized to develop a fuzzy reasoning-based predictive tool. The novelty of this study lies with the application of a clustering algorithm, which can ensure both compactness and distinctness of the developed clusters, and development of a reasoning tool utilizing the information of obtained clusters. Two types of multi-sensors data were used to test the performance of the proposed algorithm. Results were compared with those available in the literature, and the developed technique was found to perform better than the previous approaches on both the data sets. The better performance of the proposed algorithm could be due to its in-depth search of the data set through similarity-based fuzzy clustering followed by the development of fuzzy reasoning tool utilizing the information of obtained clusters.

1. Introduction

Multi-sensors data fusion deals with the extraction of useful information by combining the data sets collected by multiple sensors recording an event. A particular attribute of the event is measured by only one sensor. Each sensor may have a separate range of operation. The different attributes measured by the sensors are generally assumed to be independent of one another. Multi-sensors data fusion could be able to extract such information, which might not be possible to obtain using a single sensor. The concept of data fusion is not new. It was primarily developed for military and defence purposes, but it soon became an indispensible tool in the field of information retrieval for a wide range of applications. Hall and Llinas (1997) elaborated on the importance of data fusion for different non-military applications, such as in robotics, surveillance systems, medical diagnosis, environmental monitoring etc. The art of data fusion has evolved over the years from simple algorithms of limited applications to a plethora of sophisticated learning techniques, which could extract useful information from a variety of sensor specific data. Since the data emerge from different sensors have varying accuracies and coverage factors, benefits of data fusion include improved system reliability and/or redundancy, extended coverage, and possible shorter response time. A detailed description of the different aspects of multi-sensors data fusion and the type of datasets that can be analyzed by such methodology had been discussed in detail by Khaleghi, Khamis, Karray, and Razavi (2013). The most fundamentally challenging problem in data fusion is the analysis of imperfect data, i.e., datasets with inherent imperfections in single or multiple attribute(s). Probability theory (Dehghan Niri, Farhidzadeh, & Salamone, 2014), fuzzy set theory (Zadeh, 1983), possibility theory (Baati, Hamdani, & Alimi, 2014), rough set theory (Bazan, Nguyen, Nguyen, Synak, & Wroblewski, 2000), and Dempster–Shafer evidence theory (DSET) (Deng, Su, Wang, & Li, 2010) are the frequently used techniques to represent such imperfections and correlate different sensory data.

Multi-sensors data consists of two parts, namely input and output. The input part corresponds to the recorded values from all the sensors, which are to be analyzed; while the output part consists of already existing or predictable information that we seek to obtain by the fusion process. Such an output part is indispensable to the system, as it is always used for training the proposed fusion algorithm. Training is essential to all multi-sensors data fusion processes. In studies, where fuzzy logic is used to dissolve the available data, training is achieved by user perception of the scenario. The user decides the so-called rule base of the fuzzy reasoning tool based on his/her scientific judgement, which is then optimized as required (Zervas, Mpimpoudis, Anagnostopoulos, Sekkas, & Hadjiefthymiades, 2011). In probabilistic and possibility-based

* Corresponding author.
E-mail addresses: sagardip@umich.edu (S. Majumder), dkpra@mech.iitkgp.ac.in (D.K. Pratihar).

https://doi.org/10.1016/j.eswa.2018.04.026

approaches, the uncertainties in the dataset are exploited to account for specific output information.

In general, data fusion techniques are quite exhaustive and widely used, but they lack generality. It is extremely difficult to design systems, which are capable of analyzing different types of sensory data. In such scenarios, heuristic algorithms tend to be more effective at extracting meaningful information from the data available at hand. The aim of this study, therefore, was to develop one fuzzy reasoning tool using the concept of fuzzy clustering, which could be able to tackle a wider domain of multi-sensors data.

2. Literature review

A number of methods had been proposed over the years to fuse data collected from the multiple sensors. Carlson (1990) presented a method based on Kalman filtering to fuse data from the sensors having the same sampling rates. Kazerooni, Shabaninia, Vaziri, and Vadhva (2013) developed a federated ensemble of Kalman filter algorithm. Other popular state estimators, such as particle filters (Khan & Ramuhalli, 2009) and H-infinity filters (Hou, Jing, Gao, & Yang, 2013) had been used to fuse multi-sensors data. Soft computing methods, such as fuzzy logic (FL) (Naeem, Sutton, & Xu, 2012), genetic algorithms (GAs) (Liu, Cal, Sill, & Wang, 1996), and neural networks (NNs) (Fincher & Mix, 1990), had also been utilized for this purpose. Wavelet methods were also developed to fuse data from various sensors with different sampling rates (Chou, Willsky, & Benveniste, 1994). Probabilistic data fusion models with Gaussian distributions had been used by Luo, Lin, and Scherp (1988) for intelligent robotics applications. MRI and CT-Scan images were fused using Wavelet transform and neuro-fuzzy systems by Rajkumal, Bardhan, Akkireddl, and Munshi (2014). Deep learning algorithm had been exploited in the field of medical diagnosis to diagnose Alzheimer's disease by Suk, Lee, and Shen (2014). A major issue with probabilistic models of data fusion is their complexity in problem formulation and handling, and their relative difficulties of implementation, when compared to the techniques like fuzzy logic and neural networks. While computational power might be an issue for such fuzzy/neural systems, they remain at par with the probabilistic models, if not better at times, in terms of performance and accuracy. Due to their ease of formulation and execution, a majority of recent research in the field of multi-sensors data fusion has been carried out with such soft computing tools. Jiang, Zhang, and Zhang (2011) used fuzzy-neural techniques for determination of structural damage. Manjunatha, Verma, and Srividya (2008) achieved multi-sensors data fusion of wireless sensory data for fire detection using fuzzy logic technique.

Fault detection problem of planetary gearboxes was solved using ANFIS (Adaptive Neuro-Fuzzy Inference System) modelling by Lei, Lin, He, and Kong (2012). Forecasting of extreme weather events like flood, drought and rainstorm is generally done using a precipitation concentration index (PCI). Petković, Gocic, Trajkovic, Milovančević, and Šević (2017) developed an ANFIS model to predict PCI utilizing the input parameters, namely mean winter precipitation amount, annual total precipitation, mean summer precipitation amount, mean spring precipitation amount, mean autumn precipitation amount and mean of precipitation for the vegetation period. The mean autumn precipitation amount was found to be the most important parameter for the PCI prediction. Wind speed has a significant role on the output of a wind power system. Nikolić, Mitić, Kocić, and Petković (2017) used a neuro-fuzzy approach to identify the most important input parameter influencing the wind speed. In another study, Petković, Ćojbašić, and Nikolić (2013) designed an ANFIS model to estimate wind turbine power coefficient as the function of two inputs, namely blade pitch angle and tip-speed ratio. Moreover, an ANFIS-based intelligent controller for wind generator equipped with continuously variable transmission (CVT) was developed by Petković, Hamid, Ćojbašić & Pavlović (2014) to ensure the maximum output power of wind turbine. In a wind farm, the performance of a turbine may get affected, if it lies within the area of turbulence caused by another turbine. Therefore, it is essential to have a proper project management in order to maximize the power produced in the wind farm. Petković et al. (2014) conducted a study using an ANFIS to identify the most important input variables in order to maximize the generated power. Moreover, Petković, Nikolić, Mitić & Kocić (2017) developed a neural network-based model for prediction of wind speed fluctuation using its fractal characteristics as the inputs. In another study, a thorough investigation was carried out to use the theory of inventive problem solution in order to achieve innovative wind turbine system by Nikolić et al. (2016). An ANFIS model was also designed by Petković, Pavlović, and Ćojbašić (2016) to maximize wind farm efficiency by finding optimal layout of the turbines after considering its power production, cost per power unit and total cost. Thus, soft computing techniques specifically neuro-fuzzy systems had been widely used to extract useful information from the data sets related to nonlinear and complex systems.

Although some studies had been reported on multi-sensors data fusion using several approaches, there is a chance of developing more efficient algorithms. The objective of this study was to develop a multi-sensors data fusion algorithm using the concepts of fuzzy clustering and reasoning. A fuzzy clustering algorithm will be used, which will be able to automatically decide the number of clusters after ensuring their quality in terms of compactness and distinctness. A fuzzy reasoning tool will then be designed utilizing the developed clusters of the data set. The performance of the developed algorithm will be tested on two different data sets.

The remainder of this paper has been organized as follows: In Section 3, the proposed algorithm has been discussed in detail, while in Section 4, a brief description of the data sets used in this study is given. Results are stated and discussed in Section 5, followed by some concluding remarks in Section 6. The scopes for future studies are discussed in Section 7.

3. Proposed algorithm

The entire process of data fusion was carried out in two stages. The aim of the first stage was to generate the representatives of the dataset in hand using clustering. Clustering is an important tool in data mining in order to analyse inherent correlations in a given dataset. Its importance in multi-sensor data fusion had been elaborately discussed by Han and Lei (2012). The most commonly used conventional clustering approach is the k-means algorithm. However, in this study, a fuzzy clustering algorithm was used to take care of the inherent fuzziness of the data sets. Based on the similarities of the data points, a few clusters would be formed.

For efficient fusion in the subsequent phase, the clusters so formed must be fairly distinct. At the same time, the data points belonging to a cluster must be close enough to be efficiently represented by the corresponding cluster centre. Hence, there is a need to produce both distinct as well as compact clusters for proper retrieval of inherent information. In order to achieve the desired cluster characteristics, Entropy-based Fuzzy C-Means (FCM) clustering algorithm, as proposed by Dey, Pratihar, and Datta (2011), was used. Once the cluster centres were obtained, the second stage of the fusion system dealt with a fuzzy reasoning process to mimic input-output relationships of the data sets, as proposed by Chattopadhyay, Pratihar, and De Sarkar (2009).

An important step before clustering is to ensure similar order of magnitude for each sensory data. This is necessary, so as to render information from every sensor important. This step could be carried out by normalizing the chosen attribute data using its domain