Computers in Industry 75 (2016) 27–34

Contents lists available at ScienceDirect

Computers in Industry

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

# Multi-sensor data fusion by a hybrid methodology – A comparative study

Sarvesh Rawat [a,*], Surabhi Rawat [b]

[a] School of Electronics and Electrical Engineering (SELECT), VIT University, Vellore 632014, India
[b] School of Biotechnology & Biosciences, Lovely Faculty of Technology & Sciences, Punjab, India

ARTICLE INFO

*Article history:*
Received 4 May 2015
Received in revised form 14 October 2015
Accepted 29 October 2015
Available online 13 November 2015

*Keywords:*
Multi-sensor data fusion
Hybrid architecture
Neural network
Diagnostic tool
Rough-sets

ABSTRACT

Multi-sensor data fusion is considered as an inherent problem in wireless sensor network applications. It is widely assumed as a sturdy non linear system in view of the complexities involved in its operation. An accurate and precise methodical solution is therefore a complicated task to accomplish. It is crucial for the sensory systems that they should not be influenced in terms of accuracy and precision by any means. To address these issues a hybrid model employing rough set (RS) with back-propagation neural network (BPNN) is used to ameliorate the data fusion capability of the system with an illustrative example. Experimental results have demonstrated an escalating improvement in the predictive accuracy of the hybrid model as compared to the traditional BPNN model.



## 1. Introduction

A multi-sensor system sustains to several abrupt anomalies like electronic disturbances and thermal changes which changes the state of the system [1]. Therefore, a framework should be primed that guarantees a stable and smooth function under the influence of unreliable external or unknown influences. Systems employing multi sensors for their functions gain more informative knowledge than the one using single sensor system [2]. Complex systems used in missile tracking, live targets tracking, employs various kinds of sensors for acquiring high accuracy, improved reliability and robustness of the system [6]. Therefore, an inconceivable stress is put upon investigating the calculations focusing on multi-sensory data combinations [3]. These systems have become extremely vital these days owing to their high accuracy and large parameter strengths. There are several issues involved in a sensory network that make data fusion a complicated task. The real world data fusion application tackles with several data related challenges that are shown in Fig. 1. The input data to a multisensory fusion system is inconsistent, imperfect, correlated and/or in disparate forms. Each of these can be subcategorized further into distinctive subsets and discussed in the following section.

The greater part of the information related issues emerge from the restrictions existing in the sensor technologies and unprocessed raw data that can be categorized as follows.

*Data imperfection:* It exists when data provided by sensors is influenced by some degree of uncertainty in the measurement system or due to impreciseness present in the data. An imperfection in the data can be caused due to the ambiguity present in sensory data or due to false information value. Additionally, sometimes this imperfection also exists due to incomplete/missing information. A data fusion algorithm should express such imperfections present in the data and viably reduce them.

*Inconsistent data and data correlation:* There are several causes for a dataset to be inconsistent. When a system is encountered with outlier/spurious data or encountered by disordered or conflicting data that are primarily caused by inconsistencies and ambiguities present in the environment, it fails to distinguish between the real data and the associated ambiguities with it. The Sensory Data fusion algorithm should exploit the redundant data to alleviate such effects. Data correlation is often experienced in distributed data fusion setting. For example, a WSN has several nodes that are exposed to the same external noise, hence biasing the measurements. These data dependencies must be accounted for the exact operation of the system, in order to refrain them suffering from over/under confidence in the effect.

*Disparateness*: It alludes to the integration of the two or more disparate sources, that helps in overcoming some of the limitations

* Corresponding author.
*E-mail addresses:* sss.sarvesh888@gmail.com (S. Rawat), surabhirawat40@gmail.com (S. Rawat).

http://dx.doi.org/10.1016/j.compind.2015.10.012
0166-3615/

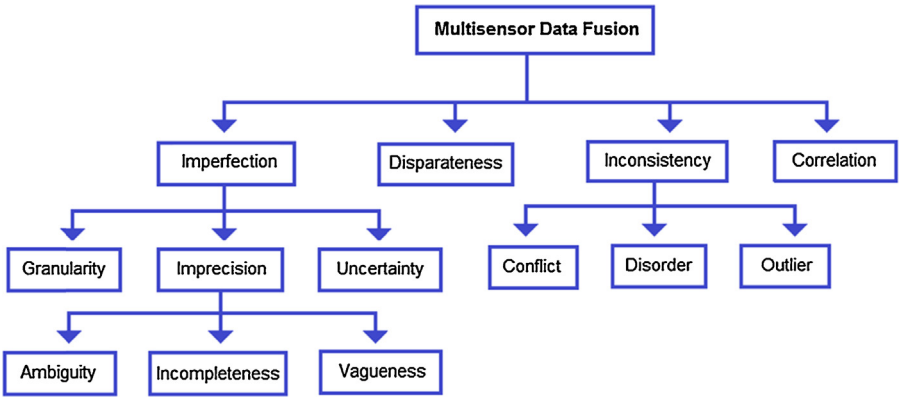


**Fig. 1.** Taxonomy of inaccuracies present in the multisensory data fusion system.

of the single source based system. While, on the other hand, a multi-sensor system comprises of different data domains which add another challenge for the data fusion [26]. For example, a typical sensory system consists of temp sensor, pressure sensor and moisture sensor. Likewise, in the same manner, some sensors deal with the chemical and physical nature of the material. Each sensor has its own signature and has its own classification and description. It becomes very hard for the system's operator to assimilate the data and mine effective rules from it [27].

The traditional sensor fusion methodologies are highly imprecise, uncertain and inaccurate since the unverifiable parameters and obscure attributes lead to a large error and complications [5,6]. The outside disturbances and uncertainty factors are very complex and non linear in nature. To analyze the whole incoming sensory data, identify the state of the system and effectively mine results from it become a difficult task for the system's operator. Hence, there is a huge demand for developing new sensor systems which can partition the error gap. To redress these raised issues several hybrid methodologies are employed such as fuzzy logic, artificial neural network, neuro-fuzzy and rough sets [11,12,16]. To identify the state of the system using the training data is a primary reason of using machine learning and soft computing techniques.

The aim of this work is to propose an effective and proficient architecture capable of delivering a high predictive accuracy. To meet the objectives, a hybrid architecture employing RS–BPNN is employed in this work to enhance the reliability and capability of the multi sensor fusion system. The proposed methodology is tested in a water pollution monitoring system. It utilizes the data, taken from the monitoring system, to train and test the proposed hybrid network. RS preprocess and mine the incoming data receiving from the sensors. It identifies the crucial features and removes the superfluous information from the database. The important features are fed to the NN classifier and trained with BP algorithm to predict the pollution level of the area.

The rest of the paper is arranged as follows. Section 2 discusses about the water pollution monitoring system. Section 3 discusses about the work done by the various researchers and the prevailed problems of this field. Sections 4 and 5 consist of the basics of RST and NN, while Section 6 shows the experiment and the results. Finally, the conclusion of the work along with discussion is presented in Section 7.

## 2. Water pollution monitoring system

Water pollution is considered as one of the most critical problems of the modern world [25]. Rapid industrialization has led to a surplus increment in the quantity of toxins and wastes in the water. Now a day, monitoring the water pollution level in water sources has become a challenging task [25]. In our example, a multisensory system containing four sensors (sensor 1, sensor 2, sensor 3 and sensor 4) is used to monitor the Biochemical Oxygen Demand (BOD) in the water area. The other four sensors (sensor 5, sensor 6, sensor 7 and sensor 8) monitor the Chemical Oxygen Demand (COD) of the water. The way of installation of the sensor net also affects the final fusion result that is represented as sensor 9. Another sensor denoted as sensor 10 that describes the status of the sewage disposal machine stating if it is running or not. Fig. 2 shows the input data aggregated from the sensors.

The output value that determines the status of water pollution level is shown in Fig. 3. The range of output values lies between 3 and 4.

The system operator analyzes all the incoming data collected from different sensors to estimate the status of the pollution level in the water. The data samples consist of multiple domains of values that make it a typical data fusion problem. There is an acute need of devising a precise and detailed mathematical model which can model the whole system. To resolve this problem, a hybrid method is proposed in this work that integrates data fusion and data mining together to produce effective rules for the system's operator.

## 3. Literature review

There are several issues associated with data fusion methodologies making it a challenging task for the system's operator as well as researchers. The majority of these issues emerged from the unprocessed imperfect data that is highly uncertain and vague in nature as discussed in previous sections. It is quite troublesome task for the system's operator to process the data from disparate sources and mine useful results. Various researchers have done a significant work to deal with the issues associated with the multisensor data fusion system.

Few researchers utilized a probabilistic approach with Bayesian network in their work to deal with the data fusion problem [19,21].This approach relies on the probabilistic density function to express the uncertainty present in the data. Although it is a traditional approach to treat the data filled with uncertainty but it is incapable of addressing the other imperfection aspects of the data, such as dealing with the missing data samples or data correlation [20]. Few other researchers used fuzzy reasoning technique to process the vague data and generate the decision rules for the system's operator [11]. An integrated system employing fuzzy with data fusion was employed to minimize the failure risk in an integrated vehicle systems [4,7]. The major limitation while using fuzzy based system lies in the fact that they are limited merely to the fusion of the vague data [15]. An