Real-time data mining methodology and emergency knowledge discovery in wireless sensor networks

Weiwei He\textsuperscript{1}, Shuanghua Yang\textsuperscript{2}, Lili Yang\textsuperscript{3}

\textsuperscript{1}Department of Computer Science, Loughborough University
\textsuperscript{2}Business School, Loughborough University
\textsuperscript{3}Loughborough, Leicestershire, LE11 3RX, UK

W.He@lboro.ac.uk
S.H.Yang@lboro.ac.uk
L.Yang@lboro.ac.uk

Abstract— Real-time data mining can extract event knowledge for fire detection more promptly than the traditional detection approach. The paper mainly describes the real-time data mining process for fulfilling rapid emergency response in wireless sensor network. We evaluate the performance of this process by simulation.

Keywords— Wireless sensor network, real-time data mining, emergency response, dynamic data process, fire detection system

I. INTRODUCTION

Wireless sensor networks have attracted significant research effort in recent years. A wireless sensor networks is a collection of autonomous nodes or terminals, which communicate with each other by forming a multi-hop radio network which is continuously connect in a decentralized manner [14]. Based on wireless sensor network, multi-sensor with intelligent methods is one of the most important developments in automatic emergency response technology. In this paper, we investigate wireless sensor networks real-time data mining dealing with multi-sensor datasets for emergency response systems in wireless sensor networks. The tasks of real-time data mining are to determine the permitted range of values of some parameters (e.g. temperature, CO, smoke, flame), to detect abnormal events (e.g. fire) and to estimate parameters of any detected event (e.g. the intensity degree of fire). The motivation is to reduce various existing problems in traditional emergency response such as the ‘early detection of fire’, ‘delay warning’, ‘a false alarm’, ‘simple management’ etc.

In this paper, we review real-time data mining in wireless sensor network and propose a case study used for emergency response. The emergency response system employs this proposed real-time data mining to improve the reduction of false alarms, which could be beneficial for the effectiveness of any determined evacuation.

In section II, we present the requirements for an emergency response system. Section III presents a detailed introduction and gives a review of recent development in real time data mining. Section IV demonstrates a case study to indicate how real-time data mining fulfill the requirements of emergency response. Section V concludes the review of existing challenges and indicates any future work.

II. EMERGENCY RESPONSE SYSTEM REQUIREMENTS

Effective emergency response systems are required to detect and correctly and quickly respond to disasters such as fires. Moreover another important requirement for emergency response, is dealing with existing data to provide forecasts of the extent to which any fire would spread. Determined from an initial study, some essential functional requirements for an effective and flexible emergency response system are listed below:

- Process and support for all event stages.
- Response to the event in the early stages.
- Provide a temporal-spatial based data.
- Be scalable to suit different environment.
- Provide support for various specific users’ requirement.
- Be accuracy and robust.
- Predict future developments before their real occurrence.

III. REAL TIME DATA MINING METHODOLOGY

Fayyad further viewed data mining to be a particulate step in large process of knowledge discovery in database (KDD) [7]. KDD is considered to be more encompassing process that including understanding application domain, data integration and selection, data mining, pattern evaluation and finally...
consolidation and use of extracted “knowledge” [8]. The Gartner Group stated that “data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technology as well as statistical and mathematical techniques”[12]. Applying this definition, to the context of emergency response systems, suggest a need to extract useful and previously unknown or unexpected emergency knowledge from a large amount of real-time environmental data in order to detect a fire incident and support emergency services to enable them to provide a more effective response. Real-time data mining combines modern data mining techniques with time series analysis techniques. The traditional data mining method focus on a limited perspective, which is inappropriate for analyzing data from an assumed model. So it cannot process time series data and newly emerged type in real time. As shown in Box and Jenkins [2] and a vast volume of time series literature, traditional time series analysis and modeling tend to be based on non-automatic and trial-and error approaches. Time series can deal with temporal sequences of datasets and forecasting data. But development of time series modeling using non-automatic approach for large amount of data would becomes impractical. Therefore, an automatic model building is necessary. Real-time data mining is developed and combines of two technologies that would make any emergency response more effective.

In the paper, real-time data mining is focused on the development of a novel approach, which aims to provide fast and effective emergency response. In emergency response, firstly, we have to collect and relay data from a wireless sensor network. The collected data are stored in database and have to be spatial-temporal structured in the pre-processing stage. Secondly, we have to build models that are robust and integrate simulation tools for event prediction and evolution. In the prediction stage, the simulation forecasts have to response faster than real time changes in the situation being modelled. Thirdly, the systems have to provide feedback to the emergency services or response system according to the results detection. Finally, the output should provide intelligent support for the users’ decision-making and development of rescue plans. Real-time data mining is the methodology that linking and realizes the functions mentioned above.

A. Problem formulation

In this stage, information and the users’ requirements have to be formulated and identified. In general, these are questions, which related to pattern and relations between data. Examples of data mining questions in emergency response include: ‘How can we characterize a fire event?’ and ‘Which groups of multi-sensor data at the current time, appear to suggest the occurrence of a fire event?’ Data mining comes into its own when there are many possible relations between large amounts of questions that have to be evaluated [9]. In this stage, an intelligible description of interesting knowledge must be present clearly. Subsequently, any discovered relations are used to make predictions about the situations occurring in current event.

According to the processing of emergency event, we the set of system problems we need to consider can be divided into three stages. Firstly, the problems are ‘Does the system run correctly?’ and ‘how can we built and update the event model?’ These two questions are used to make sure the emergency response system will run smoothly. Secondly, we need to find the characteristics of the fire event, the location and danger level of the event, and the predicted spread. Finally, we have to evaluate the results and report this knowledge for decision making.

B. Pre-processing the data

Even if all related data are collected by wireless sensor network, it will often be necessary to pre-process the data before analysis [5]. Data pre-processing generally includes the following two tasks:

- Outlier detection: Outliers are data values that are not consistent with the most other observations. The reasons for outliers’ generation are usually because of measurement errors, coding and recording errors and other environmental elements. In emergency response, outliers can seriously affect the results of fire detection and could be one of the main causes of false alarms. There are two strategies for dealing with outliers. The first is to detect and eventually remove outliers as a part of the data pre-processing stage. The second is to develop robust modelling methods that are insensitive to outliers.

- Data reduction and data projection: Data pre-processing tasks usually reduce significantly the amount of data which decreases the size of the dataset for feature selection, while data projection which alerts the representation of data. Crone et al analyse various regarding the methods applied whether parameter tuning was carried out, and which methods of data reduction and projection could be observed and present results [5].

C. Real-time data mining methods

Real-time data mining is the process of extracting interested and previously unknown or unexpected knowledge from large amount of real time data. Its correctness depends on not only logical correctness but also timing constraints. Various methods in the area of data mining have been developed in order to effectively and correctly discover knowledge. Among these methods, we review two important real-time classification and real-time clustering methods.

- Real-time data mining classification

Classification is about grouping data items into classes (categories) according to their properties (attribute values). Classification methods include, for example, decision tree,
artificial neural networks (ANN), maximum likelihood estimation (MLE), linear discriminate function (LDF), support vector machines (SVM), nearest neighbor methods and case-based reasoning (CBR) [1,9,18]. The goal of this technology is to achieve the best possible classification performance for the specific requirement [21]. Classification separates dynamic data into chunks, and then integrates the classifiers knowledge from each chunk to support the final decision. There are several classification methods, such as Sprint [17], BOAT [11].

The goal of this technology is to achieve the best possible classification performance for the specific requirement [21]. Classification separates dynamic data into chunks, and then integrates the classifiers knowledge from each chunk to support the final decision. There are several classification methods, such as Sprint [17], BOAT [11].

---

**Real-time data mining clustering**

Clustering involves the task of grouping similar objects into the same classes whereas different classes show different characteristics. Ordonez [16] has improved the k-means algorithm to cluster real-time data. The improved k-means algorithm can be applied to both real and synthetic data sets. His main idea is that it updates the cluster centers and weight after processing a batch of data rather than updating one by one. Crespoa and Weberb [4] presented a methodology for dynamic data mining using fuzzy clustering that assigns static objects to dynamic classes. Changes that they have studied are the movement, creation, and elimination of classes together with any combination of them. Chung and Mcleod proposed a mining framework that supports the identification of useful patterns based on incremental data clustering, they focused their attention on news stream mining, and they presented a sophisticated incremental hierarchical document clustering algorithm using a neighborhood search [3]. The clustering center would be changed by environmental feedback. Hence, real-time clustering method could improve the scalable of emergency response systems for various environments.

**D. Event detection**

A knowledge base is a database that stores data and presents the features of specific events. It contains and represents discovered knowledge during real-time data mining.

One of any emergency response requirements is that the emergency response systems must be suitable for different environment and users’ requirements. Accordingly, we built various models of emergency response systems. Model selection is conducted to select the relative models from a knowledge base for upcoming data analysis. In this process, as we mentioned above, it is mainly real-time classification methods that are used for choosing and classifying the various models. Many methods of model selection employ a simple approach, which prefers the use of the simplest model that fit the data [22]. For emergency response, model selection is mainly based on environment models such as global or local models. Next, the model evaluation step is used to evaluate model performance. There are two purposes of model evaluation: the prediction of how well the final model will work in the future, and as an integral part of methods which help to find the best performance for data mining. Souza et al indicate five patterns which deal with the first case for the evaluation of data mining models [19]. Finally, the last step in this stage is model updating which is to update models in the knowledge base and may fulfill dynamic models for real-time emergency response systems.

**IV. CASE STUDY**

We use a simplified example of a real-time fire detection system to illustrate how the real-time data mining process works in practice [13]. The aim of the fire detection system is to reduce false alarms, which is one of the most important and challenging application areas in emergency response systems. The requirement for this example is to improve the correctness of fire alarm detection and compare the results with the traditional fire detection system.

Data simulation is the first step for the fire detection. The sensor data include four parts: the normal data, fire data, system noise and the outlier. We assume the range of normal environmental multi-sensors data. The fire event is simulated...
and generates random fire data for each sensor. According to the different types of sensors, we assume different noise and outlier rate for each sensor.

After data preparation, the first stage of real-time data mining is to list the problem formulations. In this fire detection system, it is used to reduce false alarms. So we need to know ‘How are the false alarm generated?’ and ‘How can we improve the fire detection efficiency?’ There are various reasons for false alarm generation, such as environment changes, system noise, communication error and the outlier. In this system, we focus on system noise and the outlier that are often one of the main reasons for a false alarm.

Once we have collected the data, the next stage is data pre-processing in this stage outlier detection or removal and data reduction would be the main tasks. Data reduction may consist simply of such operations as editing, scaling, coding, sorting, collating, and summarization. Because of the raw data generated from the simulation system, data derived into simplified and corrected form. According to the emergency response requirements, the data must be based on spatial-temporal patterns. So the raw data will be integrated with time-based and spatial-based values, which indicates the data collected time and location of each sensor in the readings from the networked nodes. The advantages of the spatial-temporal patterns are to improve the information for results presentation and suitable for flexible environment. The disadvantage is that, the whole network data have to process by spatial-temporal pattern, which may extend the response time for event detection. In another tasks of the stage mentioned above, we apply a Kalman filter for outlier detection. In the system, the Kalman filter method is not only used to remove the outlier, but also to predict data in the next time stamp which could be used for knowledge updating and environment data adjusting. There are five steps in using the Kalman filter method in the system: project the estimation ahead, project error covariance ahead, calculate the Kalman gain, update the estimate with measurement which help to detect whether it is a real event or an outlier, and update the error covariance. The processing will be repeated during the operation.

In this simplified case, the environment would be simply compared with the real world. In order to analyse false alarm problem only, we integrated the model updating and selection steps and then built threshold-based model for fire detection. The threshold –based model is to apply the threshold of normal patterns of behaviour to detect emergency event. The process of threshold-based model can be shown in Fig. 1. According to real experience, we fix the threshold values for each sensor (maximum value, the increase rate) and the weight of each sensor which means the essentiality for abnormal event detection. Then repeat the processing by time label.

The simulation results show that applying the model integrated with Kalman filter and threshold-based event detection method can improve the detection reliability. From analysis, it shows the improved system could reduce by approximately 50% the number of false alarm compared to traditional system. But the real fire event can be more complex and flexible, as well as must be supplied by various models to deal with real-time environment changes in emergency response systems.

V. CONCLUSION

Real-time data mining is a developing field of study, which has raised many challenges to be addressed [6, 10, 15]. Emergency response is a highly complex process that needs the support of wireless sensor networks. Real-time data mining technologies for wireless sensor networks could greatly improve the ability of emergency response system and reduce loss of life and false alarm, as well as reducing the risks of emergency.

The existing techniques of real time data mining cannot meet the requirements of the real world. Some of them try to develop complex computational algorithms, which introduce a significant margin error in real experience. These purposed algorithms may not find out in the simulation study. For example, in our emergency response system, the environment is simpler than the real world. The fire spreads in the simulation system at a fixed speed. But in the real world, it may be affected by a changing and complex environment. Hence, in future work, we have to improve the models and algorithms to fulfill the more complex requirements from the real world.

Wang indicated [20] the scalability of data mining methods is constantly being challenged by real-time systems that generate tremendous amount of data at unprecedented rates. The fundamental problem we need to solve is as follows: “given a near infinite amount of continuous measurements”, “how do we model them in order to capture time-evolving trends and patterns in real-time”, and “make time-critical predictions”.

Real-time data mining is still in its infancy state. Further developments would be realized over the next few years in various applications, such as physical and astronomical applications [2], as well as business and financial ones etc.

REFERENCE


