Expert Systems With Applications 90 (2017) 356–373

Contents lists available at ScienceDirect

Expert Systems With Applications

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

The TUBE algorithm: Discovering trends in time series for the early detection of fuel leaks from underground storage tanks

M. Gorawski*, A. Gorawska, K. Pasterak

Institute of Informatics, Faculty of Automatic Control, Electronics, and Computer Science, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland

ARTICLE INFO

Article history:
Received 1 March 2017
Revised 7 August 2017
Accepted 8 August 2017
Available online 14 August 2017

Keywords:
Trend detection
Leak detection
Anomaly detection
Time series
Petrol station
Quality of data

ABSTRACT

Leaks and spills of hazardous fluids like petroleum endanger the environment, while remediation costs and penalties imposed when petroleum contaminates the ecosystem affect economics heavily. Therefore, it is crucial to detect any possible symptoms of a leak as soon as possible. Most of existing leak detection techniques require specialized equipment to be used, while purely software-based methods rely solely on data analysis and are very desirable since they can be deployed on petrol stations without any changes to the existing infrastructure. Moreover, such techniques can be considered as complementary to the hardware leak detection systems, as they provide additional security level. In this paper we present the TUBE algorithm, which detects fuel leaks from underground storage tanks, using only standard measurements that are normally registered on petrol stations, i.e. the amount of stored, sold, and delivered fuel. The TUBE algorithm is an autonomous solution capable of making decisions independently as well as supporting human-made decisions and thus can be considered as an expert leak detection system. The TUBE algorithm introduces a new data mining technique for trend detection and cleaning data over time series, which can be easily adapted to any other problem domain. A trend detection technique, called *tubes*, created for the TUBE algorithm is a novel data analysis method that allows to envelop uncertainties and oscillations in data and produce stable trends. Trend interpretation technique described in this paper has been designed especially for fuel leak detection purposes using our industrial experience. This paper includes a step-by-step usage example of the TUBE algorithm and its evaluation according to the United States Environmental Protection Agency requirements for leakage detection systems (the EPA SIR standard). Such an evaluation involves calculating the probability of detection and the probability of false alarm. The TUBE algorithm has obtained 98.84% probability of detection and 0.07% probability of false alarm while rejecting 42.22% of analyzed datasets due to their uncertainty. Rejecting datasets from analysis is compliant with the EPA SIR standard; however, rejection rate higher than 20% is not acceptable. Therefore we have evaluated the two-phase filtering stage of the algorithm in order to find the best combination of filters as means of data cleaning. Moreover, we have discussed the results pointing at the overall data quality problem, since it is the main cause of rejecting some datasets from the analysis. Finally, the TUBE algorithm has obtained 93.11% probability of detection and 0.73% probability of false alarm for the best combination of all parameters with 15.56% rejection rate, which is acceptable by the EPA SIR standard. The value of probability of detection is not fully compliant with the EPA SIR standard where 95% probability of detection with probability of false alarm lower than 5% is required. We have found that the requirements for the aforementioned probabilities have been completely fulfilled for datasets representing manifolded tank systems but not for single tank datasets. Such a situation was unexpected since manifolded tank systems are generally claimed to be more complex for analysis as they are in fact systems of multiple single tanks directly connected. In this paper we have also measured the time and memory complexity of the TUBE algorithm as well as discussed the issues connected to the TUBE algorithm deployment on petrol stations using our industrial experience in the topic.

* Corresponding author.
E-mail addresses: Marcin.Gorawski@polsl.pl (M. Gorawski), Anna.Gorawska@polsl.pl (A. Gorawska), Krzysztof.Pasterak@polsl.pl (K. Pasterak).

http://dx.doi.org/10.1016/j.eswa.2017.08.016

1. Introduction

Leaks and spills from storage tanks are inevitable. According to Chang and Lin (2006), 74% of reviewed accidents were connected to liquefied petroleum, while Sementelli and Simons (1997) point that from 1982 to 1994 the estimated number of leaking underground storage tanks in USA increased from 6% to 25%. In that time, the number of storage tanks also increased from 1.2 million to several million, which reveals the problem scale. According to the official website of the Environmental Protection Division of the Georgia Department of Natural Resources (The Environmental Protection Division of the Georgia Department of Natural Resources, 2016), since 1983 there were 18 599 suspected releases in Georgia only, from which almost 14 000 were confirmed. Moreover, from January 2014 until May 2016 there were 575 confirmed releases, from which 71 required clean-up or remediation (The Environmental Protection Division of the Georgia Department of Natural Resources, 2016).

Leaks can be caused by mechanical damages, environmental factors like reactive soil (Hudak, Wachal, & Hunter, 1999), or human actions including negligence and sabotage (Chang & Lin, 2006). Consequences of a fuel leak can be considered both in ecological (Jenkins, Guignet, & Walsh, 2014; Sementelli & Simons, 1997) and economical terms (Jenkins et al., 2014; Simons, Bowen, & Sementelli, 1997; Simons & Sementelli, 1997). When hazardous and flammable liquid escapes from its container, it soaks into soil and aquifers, causing a significant threat to wildlife and human beings (Rasekh & Brumbelow, 2015; Zusman, Dubnov, Barchana, & Portnov, 2012).

Leaks can occur in the tank as well as in dispensing installations, i.e. piping that connects tanks with dispensers. Detecting leaks in pipelines is a huge research area that includes usage of hardware-based (sensors), biological (detection by trained operator or animal), and software-based methods as stated in Zhang (1997). Software-based methods in terms of data analysis utilize different techniques including support vector machine learning (Chen, Ye, Lv, & Su, 2004; Mandal, Chan, & Tiwari, 2012), pattern recognition (Zhang, 1997), fault detection clustering (Murvay & Silea, 2012), fuzzy system classification (Da Silva, Morooka, Guilherme, da Fonseca, & Mendes, 2005), expert systems (Laurentys, Bomfim, Menezes, & Caminhas, 2011; Zhou, Hu, Xu, Yang, & Zhou, 2011; Zhou, Hu, Yang, Xu, & Zhou, 2009), neural networks (Arsene, Gabrys, & Al-Dabass, 2012), or machine vision with ensemble classification for the automatic detection of cracks in pipes (Wu, Liu, & He, 2015). Layouni, Hamdi, and Tahar (2017) solved the problem of detecting and locating metal-loss defects in oil and gas pipelines by using pattern detection in MFL scans with neural networks applied to predicting metal-loss defects depth. In many of the aforementioned solutions additional sensors are still required, as in Wu et al. (2015) where assessment of pipeline condition requires image data from inside of the pipe. Images are extracted from video cameras that are moving in a pipeline on robotic carries. The same applies to Lee, Rajkumar, Lo, Wan, and Isa (2013) where support vector machine classification was used on ultrasonic transducers measurements. However successful, in terms of petrol stations installation those methods are not applicable, as though they are solving mainly long range pipelines issues or require image data. An image data analysis is also widely used in a variety of event detection problems. Kubat, Holte, and Matwin (1998) also applied machine learning techniques to detection of oil spills on the sea surface using satellite radar images.

Leak detection on a petrol station can be performed by means of monitoring vapour (European Norm 13160-1, 2003), soil, or groundwater near tanks and connected piping (Morisawa & Muto, 2012; Sacile, 2007). Moreover, tanks can have built-in solutions like second walls with interstitial monitoring using pressure or vacuum systems to detect fuel outflow (European Norm 13160-1, 2003; United States Environmental Protection Agency, 2005). Single-walled tanks and their installation can also be protected by specialized fluid or vapour sensors (Martinsanz, 2015; United States Environmental Protection Agency, 2005). Although those mechanical methods are highly reliable, they can be very expensive and require major reconstruction of a petrol station when installed. Thus many vendors decide to rely solely on software-based methods, which analyze data gathered from standard sensors located on petrol stations.

Software-based leak detection methods are mainly developed in the industry, where proper certification is often required by law in order to deploy the system on petrol stations. One of the possible types of certificates that a software-based leak detection method can obtain is the EPA SIR (Statistical Inventory Reconciliation) developed by the United States Environmental Protection Agency (1990). The EPA SIR certified leak detection method consumes an input dataset consisting of one month of daily inventory records and produces a single result representing a detected leak. A result can be either qualitative (a leak is present or not) or quantitative (a leak rate). A leak detection method can obtain the EPA SIR certificate only when its probability of detection of 0.2 gallons per hour (gph) leak is not less than 95%, while the probability of false alarm does not exceed 5%. The first version of the EPA SIR standard did not include manifolded tank systems, that were introduced in 1997 by National Work Group on Leak Detection Evaluations (1996).

According to the U.S. National Work Group on Leak Detection Evaluations (NWLGDE), since 1990 at least 17 commercial systems were positively certified for the EPA SIR (National Work Group on Leak Detection Evaluations, 2017), with the last successful evaluation performed in June 2015. Due to the fact that EPA SIR certified systems are deployed in the industry, there is no documentation on how exactly they detect leakages and, as it will be described in details in Section 5, the only method of comparing our results with those systems is by performing evaluation according to the EPA SIR standard (United States Environmental Protection Agency, 1990).

Example of a non-commercial system that focuses on the EPA SIR standard is presented in Li, Shui, Luo, Chen, and Li (2011). Although the EPA SIR standard does not specify what types of measurements the input data should contain, the data specification can be defined by a certification agency performing an evaluation. For example, in Li et al. (2011) presented system requires measurements that may not be available on each and every petrol station. Therefore, the system from Li et al. (2011) is applicable only when low-range differential pressure is measured.

Another method for detecting leakage was presented in Sigut, Alayón, and Hernández (2014). The method implements the alternative standard to the EPA SIR, i.e. the EN 13160-5 standard (European Norm 13160-5, 2005) that was developed by the European Committee For Standardization (CEN). In the EN 13160-5 standard, a format of the input data is specified in details and the standard introduces an online detection evaluation that is not included in the EPA SIR standard. The EN 13160-5 standard proposes evaluation of leak detection methods via statistical inventory reconciliation; however, strict requirements on the data used in the evaluation are what mainly differ the EN 13160-5 standard from the EPA SIR. To the best of our knowledge there are only four certification agencies accredited by the CEN and none of them has performed evaluation according to the EN 13160-5 standard yet. It might be caused by small recognizability of the standard itself and extremely strict data requirements, which prolong the evaluation by few months when data must be gathered.

When a leakage detection, especially based on non-mechanical solutions, is concerned it is important to remember that not every abnormality in data indicates a leakage. Therefore, interfer-