CHERD-1930; No. of Pages 9

ARTICLE IN PRESS

CHEMICAL ENGINEERING RESEARCH AND DESIGN XXX (2015) XXX–XXX

Contents lists available at ScienceDirect

Chemical Engineering Research and Design

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

A method for pattern mining in multiple alarm flood sequences ☆

*Shiqi Lai**, *Tongwen Chen*

Department of Electrical and Computer Engineering, University of Alberta, Edmonton T6G 2V4, Canada

ARTICLE INFO

Article history:
Received 20 March 2015
Received in revised form 22 May 2015
Accepted 10 June 2015
Available online xxx

Keywords:
Alarm flood analysis
Time-stamped sequences
Multiple sequence alignment
Smith–Waterman algorithm
Industrial alarm monitoring

ABSTRACT

Alarm flood is a serious hazard for industrial processes, alarm management techniques such as delay timers and dead-bands often are incapable of suppressing alarm floods because of the existence of consequence alarms. However, this problem could be handled by alarm flood analysis, which could help find the root cause, locate badly designed part in alarm systems and predict incoming alarm floods. In this paper we propose a method for pattern mining in multiple alarm flood sequences by extending a time-stamp adapted Smith–Waterman algorithm to the case with multiple sequences, making up one of the missing steps in alarm flood analysis. The technique involves the following new elements: similarity scoring functions, a dynamic programming equation, a back tracking procedure, and an alignment generation method. A dataset from an actual petrochemical plant has been used to test the effectiveness of the proposed algorithm.

1. Introduction

Alarm systems play an important role in industrial process monitoring. Safety, quality, and efficiency of production depend heavily on alarm configurations, especially for large plants. Nowadays, advanced technologies make it much easier for a process to be monitored by a bunch of variables using new sensors and software. This convenience, however, can result in a huge number of measured variables and their corresponding configured alarms. An alarm flood, a serious hazard for industrial process monitoring, can be one of the consequences of such vast alarm configuration. During alarm floods, operators can be overwhelmed by the large amount of alarms raised in a short period of time, leading likely to improper handling of important abnormalities. EEMUA and ISA standards (EEMUA, 2013; ISA, 2009) recommend to set the threshold alarm rate for alarm floods to be 10 alarms per 10 min based on operators' normal response time (Dal Vernon et al., 1193).

Usually a large proportion of an alarm flood are nuisance alarms, of which chattering alarms form an important part. Methods such as high density alarm plots, calculation of a chattering index, delay-timers, and dead-bands can be used to visualize, quantify, and reduce such chattering alarms (Kondaveeti et al., 2010, 2013; Izadi et al., 2009). However, by applying delay-timers and dead-bands, often one cannot totally suppresses alarms during alarm floods; the remaining alarms are mainly consequence alarms, which can be caused by three reasons: (1) process state changes such as start-up and shutdown, (2) bad alarm configurations such as redundant measurements on a single process and (3) causal relationships among measured variables. In this case, alarm flood analysis are useful to reveal correlation of alarm messages and discover possible patterns in alarm flood sequences.

The discovered patterns are helpful in alarm management. For example, the first alarm message in a pattern is usually more suspicion to be the root cause of the following alarms in the sequence. The discovered patterns can also help train the operators to handle corresponding series of alarm messages more efficiently in order to prevent the overwhelming situation during alarm floods. Some badly configured parts

☆ This work was supported by an NSERC CRD project. A preliminary version of this paper was presented at the IFAC International Symposium on Advanced Control of Chemical Processes, June 7–10, 2015, Whistler, BC, Canada.

* *Corresponding author*. Tel.: +1 587 938 3239.
E-mail addresses: slai3@ualberta.ca (S. Lai), tchen@ualberta.ca (T. Chen).

http://dx.doi.org/10.1016/j.cherd.2015.06.019
0263-8762/

Please cite this article in press as: Lai, S., Chen, T., A method for pattern mining in multiple alarm flood sequences. Chem. Eng. Res. Des. (2015), http://dx.doi.org/10.1016/j.cherd.2015.06.019

in alarm systems can be revealed by the alarm flood patterns as well. Moreover, if a pattern database can be set up, it would become possible to match online alarm messages with the existing patterns, providing operators an early warning of incoming floods and their corresponding management strategies. As suggested in the ISA standard (ISA, 2009), potential dynamic alarm management could be predictive alarming, online alarm attribute modification, and online alarm suppression.

Causality analysis can also be applied once the patterns are obtained to help recover the connections between the corresponding tags in the pattern sequences and find the root cause. In (Bauer and Thornhill, 2008) a time delay estimation method was proposed to analyze causality by taking the time delay between variables as an evidence of causality. In (Duan et al., 2012), a direct transfer entropy was proposed to not only quantify the causality between process tags but also tell if the causality is direct or indirect. A transfer zero-entropy based causality analysis method was proposed in (Duan et al., 2014a) on the basis of 0-entropy and 0-information, removing the assumption of data stationarity. In (Duan et al., 2014b) methods for root cause diagnosis of plant-wide oscillations were summarized and compared, of which data-driven causality analysis as an important branch was reviewed. Alarm correlation analysis methods were also proposed in (Kondaveeti et al., 2010; Nishiguchi and Takai, 2010; Yang et al., 2012, 2013); these methods could help restore the connections between alarm tags directly without using process data.

1.1. *Current status of alarm flood pattern analysis*

Patterns can be obtained either manually or automatically. By far, expert consultation is still the way on which industrial companies rely most when haunted by alarm floods. Process knowledge such as process and instrument diagrams is usually involved in the analysis. This approach can bring the most accurate results; however the procedure is of low efficiency and high cost. Sometimes operators' experience can be used to identify patterns as well; but this becomes almost impossible for industrial complex facilities involving hundreds and thousands of alarms in floods.

The difficulties encountered in manual analysis of alarm floods could be overcome by resorting to data mining techniques. In (Kordic et al., 2008), a context based segmentation was carried out and alarm messages during the segmented periods were filtered to be the patterns. The method requires to pinpoint a target tag beforehand to set the starting point of segmentation. The authors in (Ahmed et al., 2013) used first-order Markov chains to capture alarm floods and then clustered them based on the Euclidian distance between probability matrices; pairwise alignment was then carried out based on Dynamic Time Warping (DTW) inside clusters. In (Folmer and Vogel-Heuser, 2012), a pattern growth method was applied to find patterns in alarm floods; however, as mentioned by the authors, the proposed method was sensitive to disturbances so that the pattern in sequences had to be exactly the same in order to be recognized. An approach based on Generalized Sequential Patterns (GSP) was proposed in (Cisar et al., 2010) to help discover frequent sequences in alarm floods; the algorithm was robust to disturbances and allowed some extent of order changes between alarm messages if they were raised closely. The authors in (Cheng et al., 2013) modified the Smith–Waterman algorithm to align a pair of alarm floods and obtain their similarity scores; it tolerates to disturbances and order changes as well; but is limited to pairwise usage. The algorithm proposed in this paper is an extension of (Cheng et al., 2013) to the multiple sequence case. Significant modifications in similarity scoring, dynamic programming, back tracking procedure, and alignment generation have been made to achieve the extension.

1.2. *Background of sequence pattern analysis*

Searching for patterns in sequence database has been an interesting topic for many areas, for example, frequent transaction mining in business, homology detection in biology, text matching in information, and fault detection in process control. However the approaches are usually different since the types of input data and requirements on patterns are different in different areas.

In the business area, retailers want to find the items that are usually purchased consecutively so that they can rearrange the locations of these items in their stores to facilitate sales. Approaches include Apriori-like algorithms such as (Agrawal and Srikant, 1995), which grew short frequent sequences to longer ones through multiple scans of the whole database, tree based algorithms like the FP-growth (Han et al., 2000), which did not rely on candidate generation in Apriori-like algorithms, and (Zaki, 1998) that used a vertical data format to generate patterns. Some techniques such as the use of hash tables (Shintani and Kitsuregawa, 1996) and projections (Han et al., 2000; Pei et al., 2001) were developed as well to speed up the algorithms.

In the biology area, many multiple sequence alignment algorithms were modified from the pairwise alignment methods such as (Needleman and Wunsch, 1970; Smith and Waterman, 1981; Altschul et al., 1990; Pearson and Lipman, 1988). There were basically two types of modifications: the simultaneous approach, as shown in (Johnson and Doolittle, 1986), that gave exact optimal solutions, and the progressive pairwise approach, such as (Feng and Doolittle, 1987), that gave approximate results. One popular algorithm, CLUSTAL W (Thompson et al., 1994), which was usually used for homology detection in gene sequences, was based on the second type of modification. Recently developed have been also some statistic approaches based on profile Hidden Markov Models (HMM) such as (Eddy et al., 1995). However, a suitable HMM architecture (the number of states, and how they are connected by state transactions) must usually be designed manually, as pointed out in (Eddy, 1998).

1.3. *Contribution of our work*

In this paper we propose an algorithm that extends the one in (Cheng et al., 2013) to aligning multiple alarm flood sequences case by introducing: (1) new scoring functions that are capable of describing the similarity of items in multiple time-stamped sequences, (2) a dynamic programming equation for the iterative calculation of similarity indexes of multiple alarm flood sequences, and (3) back tracking and alignment generation procedures that can be used for multiple alarm flood sequences. With this proposed algorithm, we are able to find the optimal alignment of multiple alarm flood sequences and obtain the pattern for a selected alarm flood cluster, thus making up one of the missing steps in alarm flood analysis.

Please cite this article in press as: Lai, S., Chen, T., A method for pattern mining in multiple alarm flood sequences. Chem. Eng. Res. Des. (2015), http://dx.doi.org/10.1016/j.cherd.2015.06.019