Journal of Loss Prevention in the Process Industries 41 (2016) 178–185

Contents lists available at ScienceDirect

Journal of Loss Prevention in the Process Industries

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

Association rules mining based analysis of consequential alarm sequences in chemical processes

Jia Wang, Hongguang Li*, Jingwen Huang, Chong Su

College of Information Science & Technology, Beijing University of Chemical Technology, Beijing 100029, China

ARTICLE INFO

Article history:
Received 18 June 2015
Received in revised form
17 March 2016
Accepted 17 March 2016
Available online 21 March 2016

Keywords:
Fuzzy association rules
Data mining
Correlated alarms
Alarm time sequences

ABSTRACT

In the context of industrial alarm rationalization, the analysis of consequential alarms is helpful for finding out root alarms so as to avoid alarm flooding. Motivated by this idea, this paper introduces a weighted fuzzy association rules mining approach to discovering correlated alarm sequences. Combining fuzzy sets, Apriori algorithms and alarm time series analysis, the algorithm does not search the entire item sets to find out root causes of consequential alarms. Furthermore, by transforming the association rules into fuzzy-driven causal knowledge bases and establishing the compatible fuzzy inference mechanism, a rationalized alarm topology is eventually created. Experimental results of a chemical plant show that the novel approach taking advantage of fuzzy inferences and data mining strategies is potentially effective to remove redundant alarm sequences.

1. Introduction

In chemical plants, process alarms arise when process variables go beyond the corresponding alarm thresholds. It is crucial for human operators to grasp the association relations behind the alarms so as to immediately find out the root causes of them. Due to the interactions of operational process variables, a significant disturbance may lead to an excessive number of consequent alarms. In this context, human operators usually suffer a huge amount of redundant alarms in practical engineering. However, in face of the extensive and complex alarm information, the identification of root fault causes is acknowledged as a very time-consuming task. Therefore, it is nature that the association rules mining methodology becomes increasingly important in detecting interrelations among alarms with common causes (Schleburg et al., 2013; Ahmed et al., 2013; Cheng et al., 2013; Adhitya et al., 2014; Zhu et al., 2013).

Traditionally, the correlation analysis is mainly dependent on expert knowledge. Some structural modeling tools such as signed digraphs (SDG) are helpful for abnormality propagation path searching and hence for identifying root causes of alarms. However, these knowledge based methods are usually very time consuming to implement and strongly rely on operational experiences (Yang et al., 2012). Candy and Taisne (Candy and Taisne, 2007) identified the correlated alarm sequences based on expert systems, utilizing much information of process alarms to establish diagnosis systems. In order to reduce the analysis time and find out useful information, subsequent researchers employed advanced methods for association rules mining based on alarm correlation. However, they are usually frustrated with the difficulty in getting knowledge under the changing environment. Cisar (Cisar et al., 2009) and Hostalkova (Hostalkova and Stluka, 2010) improved the Generalized Sequential Patterns (GSP) algorithm to search for frequent alarm sequences from historical alarm data. Additionally, they pointed out that the alarm trigger time is an important factor and further added temporal constraints on the algorithm. Folmer (Folmer et al., 2011; Folmer and Vogel-Heuser, 2012) set up the Automatic Alarm Data Analyzer (AADA) based on historical alarm records, which, however, is reportedly of limited robustness to disturbances.

The Apriori algorithm is widespread in extracting association rules from transaction database with frequent itemsets. However, the algorithm has to repeatedly scan the database, which results in the high cost of runtime (Li et al., 2005; Wang and He, 2006). Even though FP-growth algorithm which transforms the data into the form of tree can avoid frequent searches of candidate sets, it is difficult to directly apply it for alarm correlation analysis (Song and Rajasekaran, 2006). As we know, fuzzy sets are able to represent ambiguous meanings, which can overcome the drawbacks of traditional crisp sets only responsible for representing elements pertaining to them (Chen et al., 2012). Inspired by this idea, many

* Corresponding author.
E-mail address: lihg@mail.buct.edu.cn (H. Li).

http://dx.doi.org/10.1016/j.jlp.2016.03.022
0950-4230/

researchers have introduced approaches to mining fuzzy association rules from quantitative data (Chen et al., 2008; Palacios et al., 2012). For instance, a weighted mining approach was proposed to find fuzzy association rules from transaction data (Yue et al., 2000). However, such kinds of algorithms are rarely presented for generating alarm association rules we interested.

The main objective of this paper is to extract potential alarm association rules from quantitative databases of temporal alarm sequences. Based on alarm similarity functions, the association rules mining algorithm can exclude redundant alarms so as to identify the significant ones. Taking advantage of membership functions associated with the data, the fuzzy metrics involved can enjoy more robustness against disturbances. Alternatively, the fuzzy association rules are readily accessible for human operators in decision making.

The rest of the paper is organized as follows. Relevant work on the detailed fuzzy association rule mining algorithm is introduced in Section 2. It follows the analysis of consequential alarms based on the fuzzy causal knowledge bases along with the compatible inference mechanism in Section 3. In Section 4, the effectiveness of the method is demonstrated with an industrial case study. Conclusions and future work are given in Section 5.

2. Fuzzy association rules mining

Owing to the numerical and continuous nature of the time series data, data mining for alarm time series sequences involves two sequential stages: time series sequences discretization and association rules generation (Kaya and Alhajj, 2003; Renganathan and Smyth, 1998). Here, a sliding time window is first introduced to segment a time series into continuous subsequences. Then an Apriori-like method is used to find frequent itemsets and extract association rules from the rule set. Considering the temporal inconsistency, different alarm sequences should be specified with different weights to uncover the low frequency characteristics of relevant alarm patterns.

2.1. Fuzzy partitions

Fuzzy partitions have been widely used in pattern recognition and fuzzy reasoning. Different linguistic attributes can be partitioned by fuzzy partition methods. Specifically, a linguistic variable is characterized by a tuple denoted by (X, T_x), where x is the variable, T_x is the set of linguistic values. Owing to the ambiguous boundaries between the linguistic values, compatible membership functions are associated with the ranges of variables, which indicates the input space mapping to fuzzy membership values between [0, 1]. In regard to the alarms, linguistic descriptors such as alarm (A) and no-alarm(NA) are assigned to each variables, shown in Fig. 1.

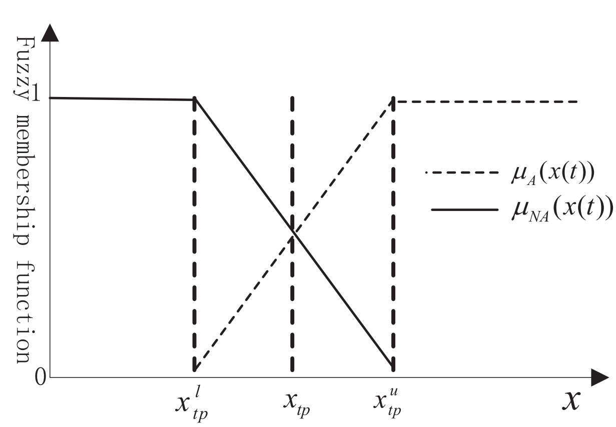

Fig. 1. The membership functions of alarm data.

In general, the following fuzzy membership functions associated with alarm variables are suggested.

$$\mu_A(x(t)) = \begin{cases} 1, x(t) \geq x_{tp}^L \\ \dfrac{x(t) - x_{tp}^L}{x_{tp}^U - x_{tp}^L}, x_{tp}^L < x(t) < x_{tp}^U \\ 0, x(t) \leq x_{tp}^L \end{cases} \quad (1)$$

$$\mu_{NA}(x(t)) = \begin{cases} 1, x(t) \leq x_{tp}^L \\ \dfrac{x(t) - x_{tp}^U}{x_{tp}^U - x_{tp}^L}, x_{tp}^L < x(t) < x_{tp}^U \\ 0, x(t) \geq x_{tp}^U \end{cases} \quad (2)$$

It is apparent that the fuzzy representation for alarm enjoys the effective treatment of data uncertainty. In the presence of the membership degrees greater than the predefined minimum supports (minsupp), the data points can be used for association rules mining. This idea can not only discover frequent alarm patterns but also decrease the numbers of false and missed alarms.

2.2. Alarm sequence discretization

A time-stamped alarm sequence is a tuple $S = (s, t_{cle}, t_{act})$, where s is the alarm tag(0,1), t_{act} is the alarm activation point, t_{cle} is the alarm clearance point. With the causal characteristics, an alarm sequence X might trigger alarm sequence Y, as shown in Fig. 2.

Definition1: Time_distance(X,Y).

Time_distance measures the time interval that an alarm sequence X triggers an alarm sequence Y, characterized by following expressions.

$$T_{x,y} = X.t_{(i)act} - Y.t_{(i)act} \quad (3)$$

A large Time_distance might involve more alarm messages associated with the alarm group, while a small Time_distance of alarms more likely implies a same root cause available. Therefore, Time_distance is considered as an important factor concerning the correlation attributes.

Definition 2: Sliding_time_window.

Sliding_time_window is used for extraction of temporal patterns and discovery of local relationships, which is computed as follows.

$$w_i = t_{(i+1)act} - t_{(i)cle} \quad (4)$$

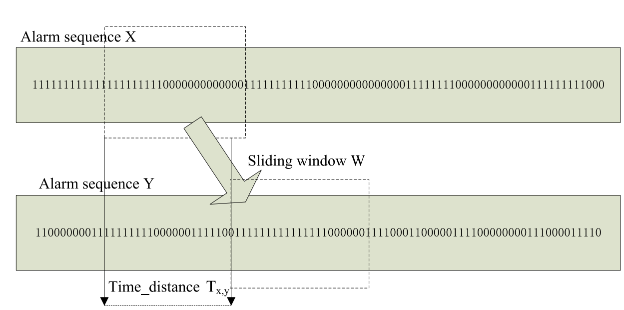

Fig. 2. Time delay of alarm sequence for interrelation rules.