Expert Systems with Applications 42 (2015) 7235–7244

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Clustering and visualization of failure modes using an evolving tree

Wui Lee Chang [a], Kai Meng Tay [a,*], Chee Peng Lim [b]

[a] Faculty of Engineering, Universiti Malaysia Sarawak, Sarawak, Malaysia
[b] Centre for Intelligent Systems Research, Deakin University, Geelong Waurn Ponds Campus, 75 Pigdons Road, Waurn Ponds, Victoria 3216, Australia

ARTICLE INFO

Article history:
Available online 7 May 2015

Keywords:
Clustering
Visualization
Failure Mode and Effect Analysis
Evolving tree
Edible bird nest industry
Euclidean similarity
Neural network

ABSTRACT

Despite the popularity of Failure Mode and Effect Analysis (FMEA) in a wide range of industries, two well-known shortcomings are the complexity of the FMEA worksheet and its intricacy of use. To the best of our knowledge, the use of computation techniques for solving the aforementioned shortcomings is limited. As such, the idea of clustering and visualization pertaining to the failure modes in FMEA is proposed in this paper. A neural network visualization model with an incremental learning feature, i.e., the evolving tree (ETree), is adopted to allow the failure modes in FMEA to be clustered and visualized as a tree structure. In addition, the ideas of risk interval and risk ordering for different groups of failure modes are proposed to allow the failure modes to be ordered, analyzed, and evaluated in groups. The main advantages of the proposed method lie in its ability to transform failure modes in a complex FMEA worksheet to a tree structure for better visualization, while maintaining the risk evaluation and ordering features. It can be applied to the conventional FMEA methodology without requiring additional information or data. A real world case study in the edible bird nest industry in Sarawak (Borneo Island) is used to evaluate the usefulness of the proposed method. The experiments show that the failure modes in FMEA can be effectively visualized through the tree structure. A discussion with FMEA users engaged in the case study indicates that such visualization is helpful in comprehending and analyzing the respective failure modes, as compared with those in an FMEA table. The resulting tree structure, together with risk interval and risk ordering, provides a quick and easily understandable framework to elucidate important information from complex FMEA forms; therefore facilitating the decision-making tasks by FMEA users. The significance of this study is twofold, *viz.*, the use of a computational visualization approach to tackling two well-known shortcomings of FMEA; and the use of ETree as an effective neural network learning paradigm to facilitate FMEA implementations. These findings aim to spearhead the potential adoption of FMEA as a useful and usable risk evaluation and management tool by the wider community.

1. Introduction

Clustering is a process of organizing a set of data attributed by multi-dimensional features into different groups based on a similarity measure (Rui & Donald, 2009). Usually, each group of data is represented by a unique weight vector, e.g. the centroid of the group (Rui & Donald, 2009). Clustering methods are useful in many applications, e.g. data mining (Lan, Frank, & Hall, 2005), data query (Lan et al., 2005), robotic arm movements (Kohonen, Simula, & Visa, 1996), noise reduction in telecommunication (Kohonen, 2001), and image segmentation (Chang, Luo, & Parker, 1998). Examples of popular clustering methods include the self-organizing map (SOM) (Vesanto & Alhoniemi, 2000), the evolving tree (ETree) (Pakkanen, Iivarinen, & Oja, 2006), fuzzy ART (Keskin & Özkan, 2009), as well as k-means (Chang et al., 1998) and fuzzy c-means (Rezaee, Leliveldt, & Reiber, 1998) clustering algorithms.

The SOM model is a neural network capable of mapping high dimensional data samples onto a lower dimensional space and representing them as nodes (Kohonen et al., 1996; Kohonen, 2001; Vesanto & Alhoniemi, 2000). It also provides a topological view of the underlying data structure (Kohonen et al., 1996; Kohonen, 2001; Vesanto & Alhoniemi, 2000). A number of enhanced SOM models have been proposed, e.g. growing SOM (GSOM) (Matharage, Alahakoon, Rajapakse, & Pin, 2011; Kuo, Wang, & Chen, 2012), growing hierarchical SOM (GHSOM) (Huang & Tsaih, 2012), and ETree (Pakkanen, Iivarinen, & Oja, 2004, 2006). These enhancements overcome two shortcomings of SOM, i.e., the requirement of a pre-defined map size before learning (Kohonen, 2001; Vesanto & Alhoniemi, 2000) and the long learning time when a large map size is initiated (Pakkanen et al., 2004, 2006). GSOM starts with a small map, and nodes are added during the

* Corresponding author.
E-mail addresses: wuileechang@gmail.com (W.L. Chang), kmtay@feng.unimas.my (K.M. Tay), chee.lim@deakin.edu.au (C.P. Lim).

http://dx.doi.org/10.1016/j.eswa.2015.04.036
0957-4174/

learning process; therefore realising a more efficient method with an incremental learning capability (Matharage et al., 2011). GHSOM has a hierarchical architecture, whereby it has a SOM-like adaptive architecture that builds various layers of hierarchy (Huang & Tsaih, 2012). For ETree, a tree structure is adopted whereby nodes are allowed to grow freely when new data samples are available (Pakkanen et al., 2004, 2006). The growing structure of ETree is important for analyzing complex data structures, e.g. in image clustering problems (Pakkanen et al., 2004, 2006). In our preliminary investigation, ETree has been shown useful for tackling textual document clustering problems (Chang, Tay, & Lim, 2013, 2014). The ETree-based approach increases the flexibility of clustering by allowing new clusters to be formed and updated, in response to new textual documents.

In this paper, the focus is on applying ETree to clustering and visualization of failure modes with the Failure Mode and Effect Analysis (FMEA) methodology. FMEA is an effective problem prevention and risk analysis methodology for defining, identifying, and eliminating failures of a system, design, process, or service (Stamatis, 2003). It has been used in a wide variety of application domains, e.g., aerospace (Bowles & Peláez, 1995), automotive (Stamatis, 2003), nuclear (Guimarães & Lapa, 2004), electronic (Zafiropoulos & Dialynas, 2005), manufacturing (Tay & Lim, 2006), chemical (Garrick, 1988), mechanical (Korayem & Iravani, 2008), healthcare and hospital (McNally, Page, & Sunderland, 1997), and agriculture (Jong, Tay, & Lim, 2013). FMEA identifies the failure modes of a system or process, understands the causes and effects of each failure mode, and determines suitable actions to eliminate or reduce the risk of the respective failure modes (Stamatis, 2003). Traditionally, the risk of a failure mode is determined by computing the Risk Priority Number (RPN) (Stamatis, 2003). The RPN model considers three factors as its inputs, i.e. Severity (S), Occurrence (O), and Detection (D), and produces an *RPN* score (i.e. multiplication of S, O, and D) as the output (Stamatis, 2003). S and O are seriousness and frequency of a failure, respectively, while D is the effectiveness of the existing measures in detecting a failure before the effect of the failure reaches the customer(s) (Stamatis, 2003).

While the effectiveness of FMEA has been demonstrated, the traditional RPN model is susceptible to a number of limitations (Bowles & Peláez, 1995; Tay & Lim, 2006; Liu, Liu, & Liu, 2013a). Many risk evaluation methods which can be used as an alternative to the traditional RPN model have been investigated. According to the review by Liu et al. (2013a), the existing risk evaluation methods can be grouped into five categories, i.e., multi-criteria decision making (MCDM) methods, mathematical programming methods, artificial intelligence methods, integrated methods, and other methods. Recently, a number of new methods for risk evaluation and/or ranking have also been developed. Bozdag, Asan, Soyer, and Serdarasan (2015) proposed an interval type-2 fuzzy set to capture both intra-personal and inter-personal uncertainties. Chang (2014) developed a soft set-based ranking technique for the prioritization of failure modes. Du, Mo, Deng, Sadiq, and Deng (2014) proposed a hybrid evidential reasoning (ER) and TOPSIS (Technique for Order Preference by Similarity to Ideal Solution)-based method for group assessment in FMEA. ER was used to express FMEA users' assessments that contain imprecision and uncertainty. TOPSIS was then used to aggregate the risk factors. Liu, Liu, and Lin (2013b) proposed a fuzzy ER and belief rule-based method for risk evaluation and ranking. The Dempster rule of combination was used to aggregate all relevant rules.

A number of new fuzzy sets related approaches have also been reported. These include an intuitionistic fuzzy set method with the weighted Euclidean distance (Liu, Liu, & Li, 2014a), a fuzzy weighted average with fuzzy decision-making method (Liu, You, Lin, & Li, 2014c), a fuzzy set theory and a multi-MOORA (multi-objective optimization by ratio analysis) method (Liu, Fan, Li, & Chen, 2014e), a hybrid intuitionistic fuzzy-TOPSIS method (Liu, You, Shan, & Shao, 2014f), and a hybrid fuzzy digraph and matrix method (Liu, Chen, You, & Li, 2014h). Besides that, Liu, You, Fan, and Lin (2014b) also proposed an D numbers and grey relational projection method, in which the assessment results were expressed in D numbers. Liu, Li, You, and Chen (2014d) presented an FMEA method comprising interval 2-tuple linguistic variables with gray relational analysis to capture FMEA users' diverse opinions. Liu, You, and You (2014g) proposed an interval 2-tuple hybrid weighted distance measure –based method for risk evaluation. In addition, a fuzzy SIRM (Single-Input-Rule-Module)-based RPN model (Jong, Tay, & Lim, 2014) and a two-stage Sugeno fuzzy-based RPN model with similarity reasoning (Jee, Tay, & Lim, 2015) were also proposed.

In our previous research, fuzzy inference systems (Tay & Lim, 2006; Jong et al., 2013, 2014; Jee et al., 2015) and an adaptive clustering method, i.e., fuzzy adaptive reasoning theory (Tay, Jong, & Lim, 2015), have been applied to FMEA. Based on the findings, we have proven the importance of maintaining the monotonicity relationship between inputs (S, O, D) and the output (*RPN* score) (Tay & Lim, 2008a, 2008b; Jee et al., 2015). Our works (Tay & Lim, 2006; Jong et al., 2013, 2014; Jee et al., 2015) focus on the development of fuzzy inference system–based frameworks for risk evaluation, with the aim of reducing the number of fuzzy rules while maintaining the monotonicity property. Fuzzy ART was used by Keskin and Özkan (2009) as a clustering method to tackle the problem whereby different combinations of S, O, and D could produce the same *RPN* scores. In addition to this reason, we further justify the advantages of using clustering methods in FMEA, as follows: (1) clustering deals with the original S, O, and D scores directly; (2) clustering allows the failure modes to be compared and visualized in the input space as groups of information; (3) the use of the original S, O, and D scores (instead of the mapped S, O, and D scores into a common domain) avoids loss of information or modification of important information for decision making purposes.

In the literature, studies on combining clustering (sometime together with visualization) and FMEA (or risk management) are not new. Arunajadai, Uder, Stone, and Tumer (2004) proposed a statistical clustering procedure to identify potential failures of FMEA. A function-failure matrix that can be used as a knowledge base to identify and analyze potential failures for new designs and redesign as well as to allow grouping of the failure modes was developed. The underlying principle was based on similarities between different failure modes pertaining to the product/component functionality. The resulting failure modes were then arranged (or visualized) in a table according to their types and clusters. The importance of grouping failure modes was again suggested by Mandal and Maiti (2014) recently. They suggested the use of a similarity to group failure modes that have similar risk levels. Besides that, Romuald Iwańkowicz and Rosochacki (2014) used a clustering method to process and analyze a database of accidents and predict the process risk. Recently, Li, Chen, and Xiang (2015) proposed a grey clustering-based indicator system to avoid the arbitrary selection of indicators in risk management of an airport safety evaluation program in China.

Studies on combining visualization and FMEA (or risk analysis and management) are also available in the literature. Inoue and Yamada (2010) visualized complicated processes in pharmaceutical research using FMEA. Grøndahl, Lund, and Stølen (2011) investigated how graphical effects (e.g., size, color, shape) and text labels introduced in the CORAS risk modeling language affected the understanding of the subject of interest. Wintle and Nicholson (2014) used Bayesian networks as graphical modeling tools for exploring what-if scenarios, visualizing systems and problems, and communicating with stakeholders during decision