Risk Prediction of Product-harm Events Using Rough Sets and Multiple Classifiers Fusion: An experimental study of listed companies in China

ABSTRACT

With the increasing of frequency and destructiveness of product-harm events, study on enterprise crisis management becomes essentially important, but little literature thoroughly explores the risk-prediction method of product-harm event. In this study, an initial index system for risk prediction was built based on the analysis of the key drivers of the product-harm event's evolution; ultimately, nine risk-forecasting indexes were obtained using rough set attribute reduction. With the four indexes of cumulative abnormal returns as the input, fuzzy clustering was used to classify the risk level of a product-harm event into four grades. In order to control the uncertainty and instability of single classifiers in risk prediction, multiple classifiers fusion was introduced and combined with self-organising data mining (SODM). Further, an SODM-based multiple classifiers fusion (SB-MCF) model was presented for the risk prediction related to a product-harm event. The experimental results based on 165 Chinese listed companies indicated that the SB-MCF model improved the average predictive accuracy and reduced variation degree simultaneously. The statistical analysis demonstrated that the SB-MCF model significantly outperformed six widely used single classification models (e.g. neural networks, support vector machine, and case-based reasoning) and other six commonly used multiple classifiers fusion methods (e.g. majority voting, Bayesian method, and genetic algorithm).

Keywords product-harm; risk prediction; multiple classifiers; self-organizing data mining; rough set

1. Introduction

Product-harm crises can be defined as 'discrete, well publicized occurrences wherein products are found to be defective or dangerous' (Dawar & Pillutla, 2000). The increasing complexity of products, the increased demands of customers, and the greater vigilance on the part of the media have made product-harm crises increasingly visible occurrences (Klein & Dawar, 2004). In America, the frequency of product-harm events in the period 2000–2004 was 1.2 times that of similar events in the 1990s (Einwiller et al., 2006). In some emerging markets such as China, rising public awareness about social rights has increased the frequency of product-harm events. Between 2006 and 2014, there were over 500 incidents related to product-harm in China, some of which involved world famous brands. Some examples are the crisis faced by Procter & Gamble in 2006 related to high levels of mercury, Sanlu's 'melamine scandal' of 2008 related to melamine-tainted milk, Shuanghui's 'clenbuterol event' of 2011 involving tainted meat products, Volkswagen's DSG gearbox failure event of 2012, and Husi Food's expired meat crisis of 2014.

Although the probability of product-harm events is relatively low for an individual enterprise, such an event, if mishandled, will not only hurt the psychology and physiology of consumers but also lead to huge economic losses and a trust crisis for the enterprise. Further, it might have undesirable social and political impacts. For instance, in an investigation conducted by Sina.com, 84% of the interviewees stated that they would no longer buy Shuanghui meat products after the 'clenbuterol' scandal, while 7.8% of the interviewees were undecided. Only 8.4% of the interviewees stated that they would continue to buy Shuanghui products, which would lead to cumulative economic losses of over 20 billion yuan for the Shuanghui group. Similarly, because of the 'melamine scandal', Sanlu Group, China's third largest dairy company, went bankrupt. Moreover, this event led to public distrust and insecurity about food safety. According to a recent market report released by AC Nielsen about China's infant milk powder, China's brand of infant milk powder had less than 20% of the market share in 2012, down from 70% in 2007.

Given that the destructiveness of a product-harm event, many challenge-seeking researchers among both academics and industry have spent a great deal of effort on the negative effects and contingency approach of product-harm crisis. However, literature shows that studies on the theory and method of product-harm risk forecasting are limited, despite the fact that they are the key processes affecting the success of crisis management. Predicting the enterprise risk is an interesting and challenging problem, and always an important concern for managers and stakeholders as well. Therefore, in this study, we attempt some exploratory research on the risk monitoring related to product-harm events. This study contributes to the literature in three ways. First, we propose a new multiple classifiers fusion method based on SODM to predict product-harm risk. Second, considering the soft boundaries of the risk grading and the main characteristics of the data such as uncertainty and vagueness, we adopt fuzzy clustering to classify risk level and divided the risk level of a product-harm event into four grades. Third, we obtain nine risk-forecasting indexes without losing information using rough set attribute reduction.

2. Literature review

2.1. Product-harm crisis

Prior case studies show that some enterprises successfully navigated product-harm crises in the international market, while some enterprises failed in the aftermath of such crises. Why is the fate of enterprises so different? There have been extensive studies on the attribution of responsibility in the context of a product-harm crisis, the impact of such crises on the relevant variables, and the coping style of the affected enterprises. The attribution results of a product-harm crisis will affect the recognised risk and responsibility judgement of the consumers, subsequently affecting their buying behaviour. Therefore, some studies examined the attribution of responsibility from the perspectives of enterprise reputation, corporate social responsibility, essential features of events and

consumers traits, etc. (Klein & Dawar, 2004; Whelan & Dawar, 2014; Carvalho et al., 2014). In order to identify the negative impacts that a product-harm event could have and to determine how these impacts could be reduced, some studies examined the impact of product-harm crises based on Eindruckstheorie, the expectancy theory, etc. These studies reported that product-harm crises mainly affected the consumers' risk perception, complaints, loyalty, purchase intention, and brand equity (Vassilikopoulou et al., 2009; Ma et al., 2014; Liu & Shankar, 2015). After a product-harm crisis, the company involved must take remedial action to change consumer perception and behaviour and to retain them at pre-crisis levels. In order to address this problem, scholars have proposed different strategies such as blame acknowledgement, voluntary product recall, and emphatic denial. The effectiveness of these strategies have been verified through experimental studies (Avnet & Laufer, 2014; Cleeren, 2014).

The review of the extent literature shows that the research on product-harm risk prediction has not proposed sufficiently valuable solutions. Through product-harm risk prediction, enterprises could be warned before the crisis; moreover, they could take reasonable coping actions according to the different risk levels. If an enterprise cannot overestimate well the risk levels of a product-harm event, it could easily lead to over-reaction or under-reaction (Rhee & Haunschild, 2006). For instance, in 2004, there were media reports that the hydrogen peroxide content in Beijing Juneng Group's calcium product could cause cancer. In fact, the hydrogen peroxide content in their calcium product was well below the legal limit. The company launched a series of measures, ranging from refunding money to apologising to their customers. However, this move did not alleviate the consumers' risk perception; instead, the consumers perceived their products to be really harmful, eventually leading to the company's bankruptcy. In contrast, during the clenbuterol event of 2011, due to underestimation of public aggressive behaviour risk, Shuanghui Group' crisis-related public relations efforts involved suppliers and distributors, but not the consumers. This action made matters worse. As another example, KFC faced the 'Sudan Red' crisis in 2005. Because of the proper risk assessment and effective actions, this event did not affect their sales; instead, the event helped to consolidate KFC's world famous brand image in the minds of Chinese consumers. Above all, it is necessary to study the prediction of product-harm risk in a systematic manner.

2.2. Business risk prediction

Since the 1960s, many studies have focused on enhancing the accuracy of business failure prediction. Prior studies usually approached bankruptcy prediction using various statistical methods such as the probit model (Lennox, 1999), the logit model (Hu & Sathye, 2015), and so on. However, these statistical methods require the predictive variables and functions to be independent. In practical terms, none of the predictive variables or functions are completely independent. In recent years, artificial intelligence models such as artificial neural networks, association rules mining, genetic programming models, case-based reasoning, and support vector

machines (SVMs) have been regarded as alternate classification technologies that can be used instead of statistical modelling to develop business failure prediction models (Kumar & Ravi, 2007; Ravisankar & Ravi, 2010; Tsai & Hsu, 2013; Geng et al., 2015). In particular, artificial intelligence techniques have been shown to have superior performance compared to statistical techniques (Tsai, 2008; Sun et al., 2015).

The extant studies on the prediction of enterprise risk suffer from a few limitations. The first issue is related to the selection of the prediction indicators. The extant research on prediction indicators is mainly based on the daily operating conditions of enterprises. Since the financial data of enterprises can be obtained from their annual reports, these studies emphasise the use of financial indicators to build their model (Wang, et al., 2015; Geng et al., 2015). However, product-harm events have obvious characteristics of an emergency; further, the root causes for the escalation of an event and the plights of the business are often not financial factors. Excluding these factors could result in an incomplete prediction model, which makes it required to study the prediction indicators of product-harm events. Secondly, the extant prediction models are mostly based on a single classifier. However, a single classifier's performance with regard to specific issues largely depends on the mode characteristics of the sample, and every single classifier has its own unique uncertainty. Multi-classifiers can reduce the variance in estimation error and improve the overall performance of the model. Therefore, a few researchers have spent a great deal of effort on multiple classifiers fusion method. But the existing fusion algorithms tend to ensemble all of classifiers at hand. In fact, in terms of prediction accuracy, a sub-ensemble of select classifiers would be superior to an ensemble that includes all the classifiers (El-Melegy & Ahmed, 2007). As a consequence, in order to reduce the one-sidedness and uncertainty of a single classifier-based prediction model, it is necessary to develop a multi-classifier fusion mechanism to forecast the risk of product-harm events.

3. Framework and Methodology

3.1 Framework

Given the limitations of the extant research, we proposed a forecasting model for product-harm events using rough sets and the fusion of multiple classifiers. Firstly, the risk level of a product-harm event was classified using the fuzzy clustering method. Secondly, based on the analysis of the key drivers of the product-harm event's evolution process, we built an initial index system of risk prediction, and then obtain the ultimate index system using rough set theory. Lastly, we selected the basic classifiers based on individual optimisation and diversity criteria; subsequently, we employed SODM technology to make a decision-level fusion on the outputs of the basic classifiers. Thus, we obtained the fusion model that satisfies the optimal complexity theory and risk level prediction. The proposed framework is presented in Figure 1.

Here insert Figure 1

3.2. Risk grading of product-harm event

3.2.1. Selection of clustering variable

A negative event can have an adverse-even disastrous-impact on the share price of a listed company. Therefore, the risk level of the product-harm event can be measured in terms of the variation in stock returns. Event studies usually use abnormal returns (AR) to study the variation in stock prices and returns in response to the related event. Brown and Warner (1985) introduced three ways to compute AR: mean adjusted returns, market adjusted returns, and ordinary least squares (OLS) market model. While all of these methods have been used in prior studies, the method based on market adjusted returns has been found to be more convenient and effective than the other two methods (Brown & Warner, 1985; Chen et al., 2013; Harris et al., 2014). After comprehensive consideration of the options, we chose cumulative abnormal returns (CAR) as the cluster variable for grading the risk after the event, and we used market adjusted returns to compute AR. Following the main findings reported in the extant research (Liu et al., 1990; Liu & Zhang, 2012), we studied the cumulative abnormal returns from the first day of exposure (R) over the next six days (R+6); i.e. the event window is [R, R+6].

According to the formula for market adjusted returns, daily abnormal returns (AR) is calculated as

$$AR_{it} = R_{it} - R_{mt}, \quad t = R, \cdots, R+6 \tag{1}$$

where AR_{it} is the abnormal returns of stock i on day t; R_{it} is the real returns of stock i on day t; and R_{mt} is the market returns on day t.

The cumulative abnormal returns of stock i from R to R+6 (CAR_{it}) is calculated as

$$CAR_{it} = \sum_{j=R}^{R+6} AR_{ij} \tag{2}$$

According to the China Security Market Accounting Research (CSMAR) database, there are four main ways to measure R_{mt}: the average returns of the whole stock market, the average returns of the stocks of the same industry, the average returns of stocks with similar size, and the average returns of stocks with similar levels of risk. Therefore, CAR, which we chose as the clustering variable for risk grading, should have four measures: the cumulative abnormal returns relative to the whole stock market (CAR_{am}); the cumulative abnormal returns relative to the same industry (CAR_{sm}); the cumulative abnormal returns relative to similar-sized enterprises (CAR_{ss}); and the cumulative abnormal returns relative to the stocks with similar risk levels (CAR_{sr}).

3.2.2. Main step of fuzzy clustering analysis

According to Boreiko (2003) and Amirian et al. (2015), the main steps of fuzzy clustering analysis are:

Step 1: Data normalisation. Different data has different dimensions. Hence, it is necessary to apply a dimensionless method to the data and compress the data to the interval [0, 1].

Step 2: Establishing fuzzy similarity relation R. Establishing the fuzzy similarity matrix is also known as

calibration. It is important to define the statistics r_{ij} that can be used to measure the degrees of similarity of the objects. If the parameter c is properly chosen, the result will be $0 \le r_{ij} \le 1$ (Meng et al., 2014). In this paper, through the calculation and adjustment, we take $c = 0.1$, which can make the calculation results in the interval $[0,1]$ to disperse better.

$$r_{ij} = 1 - c \bullet d\left(x_i, x_j\right) \tag{3}$$

where $d\left(x_i, x_j\right)$ is Euclidean distance, $d(x_i, x_j) = \sqrt{\sum (x_i - x_j)^2}$.

Step 3: Solving the transitive closure. The transitive closure of a fuzzy similarity relation R, also known as a fuzzy equivalence relation, is defined as the relation that is transitive, contains R and has the smallest possible membership grades (Yue et al., 2015). In this study, we solved the transitive closure using quadratic programming. According to $R \to R^2 \to R^4 \to ... \to R^{2^k} \to ...$, after the limited operation, there will be $R^{2^k} = (R^{2^k})^2$, and a transitive closure matrix can be calculated out $\tilde{R} = R^{2^k}$.

Step 4: Solving the cut matrix of the fuzzy matrix. Set $\tilde{R} = \left(r_{ij}\right)_{mn}$ as the fuzzy equivalence matrix. For arbitrary $\lambda \in [0,1]$, $\tilde{R}_\lambda = \left(r_{ij}^{(\lambda)}\right)_{mn}$ is the λ- cut matrix of $\tilde{R} = \left(r_{ij}\right)_{mn}$ and $r_{ij}^{(\lambda)}$ is calculated as:

$$r_{ij}^{(\lambda)} = \begin{cases} 1, & r_{ij} \ge \lambda \\ 0 & r_{ij} < \lambda \end{cases} \tag{4}$$

Solving the cut matrix when λ is equal to a certain value and clustering the corresponding rows of the matrix into the same class, we obtain the classification of the objects. λ denotes the interception level which is also known as threshold. Different values of λ are correspond with different levels of classification. According to the results of $\tilde{R}_\lambda$, the larger the value of λ, the more the number of the categories. In this paper, we decrease the value of λ gradually from 1 to 0 and find that when λ=0.83, risk grading performance is optimal and the samples could be divided into four categories.

3.3. Risk-prediction index system for product-harm events

3.3.1. Designing the primary risk prediction indexes

Determining the key drivers of the evolution process of a product-harm event is important for identifying the risk factors and building the risk prediction index system. The analysis of a large number of cases (such as KFC's 'Sudan Red' crisis, Volkswagen's DSG gearbox failure event, Husi Food's expired meat crisis, and so on) indicated that the direction and velocity of the evolution of a product-harm event are mainly affected by the initial characteristics of the event, enterprise emergency behaviour, public behaviour, adaptability of enterprise operation system, and the interactions among these factors, as shown in Figure 2.

Here insert Figure 2

The initial characteristics of a product-harm event, including the sensitivity and destructiveness of the issue,

corporate reputation, and media influence, are decisive forces with regard to the direction and velocity of its evolution (Zhao et al., 2012; Claeys & Cauberghe, 2015). According to the theory of attention, negative press about a company that is a household name is more likely to cause a 'sensational effect' (Rhee & Haunschild, 2006). Further, the authority and reliability of the source of negative information can have an impact on the credibility of rumour. The higher the authority and the reliability of the media, the better it can guide public opinion (Hovland & Weiss, 1951).

Enterprise emergency behaviour plays a decisive role in the development of the events (Sweeny, 2008), and the cognitive bias related to the risk of product-harm events is an important reason why enterprises to take improper countermeasures. The stress levels of an enterprise under normal conditions, its risk awareness, and its level of emergency preparedness can affect the policymakers' risk perception (Sweeny, 2008; Wang & Ritchie, 2013; Walumbwa et al., 2014). Additionally, public aggressive behaviour is another important factor leading to deterioration of the events. Their attitude and conduct depend on their risk perception, which is mainly affected by the sensitivity of the matter, the credibility of the media, the frequency of media reports, the extent of public involvement in the event, etc. (Park & Sohn, 2013; Goodwin & Sun, 2013; Zhao et al., 2013).

The adaptability of enterprise operation system is significant for a company's development. According to the sociotechnical system theory (Eason, 2014), when faced with negative information, a company ought to make use of all its resources, adjust the relation between internal society and technical factors, as well as the company and its stakeholders. Thus, it can tackle the external changes effectively. Some prior studies showed that the applicability of enterprise operation system in the event of a product-harm crisis is affected by the relationship between the company and its stakeholders, such as the government, suppliers, retailers, and employees, as well as by its short-time repayment abilities (Bonardi & Keim, 2005; Park et al., 2014; Meintjes & Grobler, 2014).

Based on the analysis of the key drivers of a product-harm event's evolution process, we built the initial index system for risk perception (as shown in Table 1).

Here insert Table 1

3.3.2. Simplification of the index system using rough set theory

Considering all the index data as input may not only increase the complexity of the model but also reduce the accuracy of the model because of indicator redundancy. Therefore, the rough set theory is employed to reduce the attributes of the samples. Rough set is an effective mathematical method for dealing with vagueness and uncertainty, which definitely seems suitable for the expression of vagueness and the induction of uncertainty. As it is known to all, the decision rule induction is an important work in data mining, the core idea of rough set is that knowledge reduction occurs through constant classification, and then the decision rule can be inferred from the

reduction result (Pawlak & Skowron, 2007; Zhang et al., 2003).

Discretisation of continuous attributes. Considering that the variables that were selected for the risk prediction of product-harm events are mostly continuous attributes, and since the rough set cannot deal with continuous data directly, it became necessary to adopt the clustering algorithm to discretize the continuous attributes. However, in the existing clustering algorithm, the shutdown conditions and the parameter input are all human-controlled. That is, users need considerable knowledge of the field for using the clustering algorithm (Saha et al., 2007). Further, it is difficult to ensure the quality of clustering results. Therefore, we created an objective clustering standard on the basis of information entropy theory. Further, we combined fuzzy clustering, used membership as the basis of the information entropy calculation, and applied the pedigree method to determine the number of clusters. In the discretising process of continuous attributes, in order to avoid the low search efficiency for the cluster number caused by the large index system of candidate variables, this paper adopted a heuristic searching on the basis of the maximum discernibility value (MDV) function proposed by Wang et al. (2015). Please refer to the supplementary file or the paper by Wang et al. (2015) for the detailed fuzzy clustering discretisation steps based on the MDV function and information entropy.

Attribute reduction. As the core content of rough set theory, the purpose of attribute reduction is to get the minimal attribute set without changing the resolution capability of the original information system. However, it has been proved that this problem is NP-hard (Pawlak & Skowron, 2007; Wang et al., 2015). So, it is necessary to simplify the calculation using heuristic information. The heuristic reduction method based on attribute significance has been widely researched, such as the traditional computing method of attribute significance based on algebra theory, the algorithm based on differential matrix, and the algorithm based on information theory. In this paper, we adopted the algorithm of attribute reduction based on the mutual information of binary channel proposed by Wang et al. (2015), and used it to simplify the index system for risk prediction of a product-harm event. Please refer to the supplementary file or the paper by Wang et al. (2015) for the detailed algorithm of the attribute reduction based on the channel capacity.

3.4. SODM-based multiple classifiers fusion

As noted by Feng and Wang (2015), the core element of multiple classifiers fusion is a fusion algorithm. The common fusion algorithms include Random Forests (Breiman, 2001), majority voting (Ruta & Gabrys, 2005), Bayesian approach (Kurzynski & Wozniak, 2012; Woźniak et al., 2014), and genetic algorithm (GA) (Kim & Kang, 2012), which tend to ensemble all the classifiers at hand. In fact, each basic classifier makes decisions according to the same classification task, which may lead to decision redundancy. It has been suggested that an ensemble composed of all the basic classifiers would not necessarily improve classification accuracy, while a

sub-ensemble of select classifiers may have better prediction performance (El-Melegy & Ahmed, 2007). Therefore, the key to a successful multi-classifier fusion algorithm is to fuse the classification results of multiple classifiers that efficiently conflict one another. Self-organising data mining (SODM) is a good tool for addressing this problem (Ivakhnenko, 1970; Costea, 2013).

The core technology of self-organising data mining (SODM) is the group method of data handling (GMDH). The GMDH partitions the entire sample into three parts for training, checking, and testing; subsequently, it builds the general relationships between the system's inputs and outputs. This can be expressed using a complicated discrete form of the K-G polynomial. Take the items of the K-G polynomial as the initial neurons, different pairs of which are constructed and then fed to the network's first layer. Consequently, the outputs of the neurons on the first layer are selected according to the external criterion as the inputs of the next layer. This process is repeated continuously to generate the competition models continuously until we find the optimal model. The main characteristic of the GMDH algorithm is that it can group the data and utilise both the internal and external criteria throughout the modelling process. It creates candidate neurons on the training subset according to the internal criterion and subsequently removes undesired neurons on the testing subset according to the external criterion. As the complexity of the model increases, it will reach the minimum value of the external criterion that is characterised by 'complementarity'. When the algorithm converges to a global minimum, it means that the optimal complexity model has been found, according to the optimal complexity theory. In this process, the modeller provides only the data, the corresponding data group, and the external criterion; therefore, the structure of the model is generated by the computer automatically, the influence of the subjective evaluation of the modeller on the model is eliminated, and the objectivity of the model is ensured. Therefore, the GMDH network is relatively superior to the other models (Sheikholeslami et al., 2014).

In this study, we introduced multi-layered SODM neural networks into multi-classifier fusion and built an SODM-based multiple classifiers fusion (SB-MCF) algorithm. The basic flow of SB-MCF algorithm is shown in Figure 3.

Here insert Figure 3

Step 1: The sample set is divided into training set A, checking set B, and testing set C. Let $W = A \cup B$.

Step 2: Build the general relationships between the model's output classes Y and k classifiers' classification (R_1, R_2, ..., R_k) in W. This can be expressed by a complicated discrete form of the K-G polynomial:

$$Y = f(R_1, R_2, \cdots, R_k) = a_1 R_1 + a_2 R_2 + \cdots + a_k R_k \quad (5)$$

where a_i and a_j are the weighting coefficients calculated with least squares (LS) in the training set (Lin, 2012).

By the quadratic sub-expression in Eq. (6) for each row of k data triples, the following matrix equation can be given as $\boldsymbol{R}\mathbf{a} = \boldsymbol{Y}$, where $\mathbf{a}$ is the vector of unknown coefficients of quadratic polynomial in Eq. (5), and $\boldsymbol{Y}$ is the vector of the output's value from observation. The LS obtains the solution of the equations in the form of

$$\mathbf{a} = \left(\boldsymbol{R}^T \boldsymbol{R}\right)^{-1} \boldsymbol{R}^T \boldsymbol{Y} \tag{6}$$

which determines the vector of the best coefficients of Eq. (6) for the whole set of k data triples.

Take the items of Eq. (6) as the k initial neurons fed to the GMDH multi-layered neural networks: $v_1 = a_1 R_1,\ v_2 = a_2 R_2, \cdots,\ v_k = a_k R_k$.

Step 3: Select the minimum bias criterion as the stopping criterion of the fusion algorithm:

$$\eta_{bs}^2(W) = \left\| y_t^m(A) - y_t^m(B) \right\|_{t \in w}^2 \tag{7}$$

where $y_t^m(A)$ stands for the t^{th} output value estimated by the model on dataset A, and $y_t^m(B)$ stands for the t^{th} output value estimated by the model on dataset B.

Step 4: Different pairs of the initial neurons are constructed and fed to the network's first layer; subsequently, we get C_k^2 candidate neurons of the quadratic polynomial, which is written as:

$$z_t = a_i v_i + a_j v_j \left(i,\ j = 1,2,\cdots,\ k;\ i \neq j;\ t = 1,2,\cdots,\ C_k^2\right) \tag{8}$$

Step 5: Based on the stopping criterion in Eq. (7), evaluate the candidate neurons on checking set B; subsequently, select several candidate neurons with minimum values as the inputs of the second layer.

Step 6: Repeat Step 4 and Step 5; the second layer, the third layer, … candidate neurons are generated, until the model reaches the minimum value of the external criterion, and the optimal complexity fusion model is found according to the optimal complexity theory.

Step 7: Input the classifiers' classification $\left(r_1, r_2, \cdots, r_k\right)$ in testing set C to the optimal complexity fusion model to obtain the final predication results.

4. Experiments design

4.1. Samples and data

We selected companies listed in the Shanghai and Shenzhen securities exchange that had faced product-harm crises during January 2006 to March 2015 as the samples for this study. First, we collected the samples of product-harm events from the financial investigation centre of Sina.com. Subsequently, we included the samples from Sohu.com and NetEase.com that had not been investigated by Sina.com. Sina.com, Sohu.com, and NetEase.com are China's three major portals; their finance channels provide users with the latest news, full and timely financial reports, and online surveys on hot topics. Their investigation centres involve active user participation. Finally, in order to ensure the availability of data, we eliminated the samples that were not listed in the Shanghai and Shenzhen securities exchange and those for which we were unable to ascertain the first exposure

time of the event. Additionally, companies whose predictive indexes and cluster variable data were not complete were not considered. The final sample size was 165. All the selected samples of different risk levels met the stringent matching requirements for industry and asset size. Table 2 presents the relevant information about the sample selection process.

Here insert Table 2

In this study, the data of the clustering variables was obtained from the CSMAR database. The measuring methods of each index are shown in Table 3. The data pertaining to I_1, I_{11} was obtained from the Baidu Index (index.baidu.com); the I_3, I_9, I_{10} data was obtained from the investment centres of Sina (survey.news.sina.com.cn), Sohu (survey.sohu.com), and NetEase (money.163.com); the I_2 data was obtained from the Chinese Media Influence Ranking List for 2006–2014; the I_5, I_6, I_7, I_8 data was prepared by the author based on the corporate annual reports and clarification announcements that were obtained from the Wind Financial database; and the I_{13}, I_{14} data was obtained from the Chinese listed companies' social responsibility reports that were released by Hexun.com. All the other data was obtained from the CSMAR database.

4.2. Risk grading

We used CAR_{am}, CAR_{sm}, CAR_{ss}, and CAR_{sr} as the input variables for the fuzzy cluster analysis, and used the Matlab 2014b toolbox and its associated programming language to realise the experimental process. The results showed that when λ=0.83, the samples could be divided into four categories. The risk level of the four categories could be determined in terms of the descriptive statistics of the four variables in each category. Table 3 presents the descriptive statistics and the analysis of variance of the variables after classification. The statistics shows that the sample size of giant warning, heavy warning, medium warning, and light warning were 38, 47, 45, and 35, respectively. Each variable in the different categories had significant differences (*p*-values close to zero). This means that the risk grading is reasonable.

Here insert Table 3

4.3. Reduction of indicators

The discretisation results in Table 4 were obtained based on the discretising interval adjustments according to expert advice. These experts include one professor in the field of crisis management, one professor in the field of data mining, and two operating managers of the enterprise. They fine-tuned the results of discretisation by group discussion and reached a consensus ultimately. The attribute reduction of the discretised data was implemented using rough sets, resulting in 36 reductions, including one smallest condition attribute set: {I_1, I_2, I_4, I_6, I_9, I_{11}, I_{12}, I_{13}, I_{15}}. Subsequently, we constructed the experimental dataset based on the original data of the index on the smallest condition attribute sets, and we tested the prediction performance of the model.

Here insert Table 4

4.4. Construct experimental datasets

An SODM-based multiple classifiers fusion (SB-MFC) model can improve the prediction performance and decrease the uncertainty of the prediction process. For an empirical experiment, multiple experimental datasets are needed to get multiple performance statistics. In this study, the random multiple hold-out strategy was adopted. We simulated multiple experimental datasets by repeatedly and randomly dividing the initial sample into a training sample and a testing sample. Each time, by simple random sampling without replacement, two-thirds of the companies in each of the four categories (giant warning, heavy warning, medium warning, and light warning) were selected to form the training sample; the remaining one-third of the companies in each of the four categories were used to form the testing sample. After 20 iterations of random sampling, 20 experimental datasets were generated.

4.5. Selection of basic classifiers

The performance of multi-classifier fusion directly depends on the individual recognition performance and the diversity of the basic classifiers (Sun & Li, 2008b). Therefore, in this study, we followed the diversity principle and the individual optimised principle to select the basic classifiers.

The diversity principle is based on the fundamental assumption that a perfect classifier never exists. Any single classifier is not a perfect classifier, they are required to remain certain degree of diversity. Thus, when some basic classifiers output the wrong predicted labels, the other basic classifiers could possibly output the right predicted label (Kuncheva, 2005). In order to utilise the superiority of different learning algorithms, the basic classifiers are trained using different learning algorithms (with heterogeneous model representations) on a single dataset. In this study, Q statistics was used as the diversity measure (Yule, 1900; Yang, 2011). We defined $M_{k_1k_2}^{11}$ as the number of samples that are accurately recognised by classifier f_{k_1} and classifier f_{k_2}; $M_{k_1k_2}^{10}$ as the number of samples that are recognised accurately by classifier f_{k_1} and misclassified by classifier f_{k_2}; $M_{k_1k_2}^{1*}$ as the number of samples that are recognised accurately by classifier f_{k_1}; $M_{k_1k_2}^{0*}$ as the number of samples that are misclassified by classifier f_{k_1}, and so on. The measure can be simply calculated by:

$$Q_{k_1k_2} = \frac{M_{k_1k_2}^{11} M_{k_1k_2}^{00} - M_{k_1k_2}^{01} M_{k_1k_2}^{10}}{M_{k_1k_2}^{11} M_{k_1k_2}^{00} + M_{k_1k_2}^{01} M_{k_1k_2}^{10}} \tag{9}$$

where the value range of $Q_{k_1k_2}$ is [-1, 1]; $Q_{k_1k_2}$=1 means there is a completely positive correlation between classifier f_{k_1} and classifier f_{k_2}; and $Q_{k_1k_2}$=-1 means there is a completely negative correlation between classifier f_{k_1} and classifier f_{k_2}.

Individual optimisation is based on the fundamental assumption that we should pursue a perfect classifier continuously. Although it would be impossible to create a perfect classifier, we can get a near-to-optimal classifier by optimising the model parameters or model structure. Since different basic classifiers are trained using different learning algorithms on a single dataset, the product-harm risk prediction should be based on cross-validation accuracy. Therefore, in order to optimise the performance of a basic classifier, we can optimise the parameter and the model structure through grid search and genetic search.

Based on the principles of individual optimisation and diversity, multiple discriminant analysis (MDA) (Meyer & Pifer, 1970), logistic regression (Logit) (So & Kuhfeld, 1995), back propagation neural network (NN) (Basheer & Hajmeer, 2000), decision tree (DT) (De Mántaras, 1991), support vector machine (SVM) (Weston & Watkins, 1998), and case-based reasoning (CBR) (Tseng & Chang, 2005) were selected as the six single classifiers. The detailed algorithms of the six single classifiers above are shown as Appendix A. In order to obtain the basic classifiers that were used for the multiple classifiers fusion, we optimised the learning algorithms of the six single classifiers on the same training samples. The diversity measures between each pair of basic classifiers are shown in Table 5. The statistics shows that there were no pairs of six basic classifiers with completely positive correlation or completely coincident recognition results. This means that each pair of basic classifiers was diverse to a certain degree.

Here insert Table 5

5 Experimental results and discussion

5.1. Comparison of prediction performance between SB-MCF model and single classifiers

In order to test the performance of the SODM-based multiple classifiers fusion (SB-MCF) model with regard to product-harm risk prediction, we compared the prediction performance of the six single classifier models that were used in the classifier fusion in this study; these models are widely used in the literature. The training accuracy and testing accuracy on 20 experimental datasets, and the corresponding mean, variances, and discrete coefficients of the basic classifiers and SB-MCF are listed in Table 6. The experimental results show that the average training and testing accuracy of the SB-MCF model are highest compared to the risk prediction methods based on single classifiers; moreover, the variance and discrete coefficient of the former model are the lowest. In addition, we created 20 experimental datasets based on the original data of the 17 initial indicators. Each model was trained and tested on these datasets. The results are presented in Table 6. Table 6 shows that the training and forecasting performances of each model that adopted the nine indicators after reduction as inputs were superior to the performances of the models that adopted the 17 initial indicators as inputs. This finding verified the necessity of reducing the indicators and the rationality of the reduction results.

Here insert Table 6

To examine whether the performance of the SB-MCF model was significantly superior compared to that of each single classifier model in terms of product-harm risk prediction from a statistical perspective, the difference in accuracy and discrete degree of the models on the 20 experimental datasets were analysed using the *t*-test. The results are presented in Table 7 and Table 8.

As shown in Table 7, the training accuracy of the SB-MCF model is significantly superior to that of the single classifiers at the 1% level, except the NN classifier, which is not significantly different from the SB-MCF model. The SB-MCF is significantly superior to the single classifiers in terms of the discrete degree, except the MDA and SVM, which are not significantly different from the SB-MCF model. These results are consistent with the characteristics and the experiment performance of each single classifier for two reasons. (1) There is consensus that NN has a very strong nonlinear fitting ability. As shown in Table 6, NN has the highest training accuracy. (2) The SVM and MDA are two kinds of single classifiers with the best training stability (Hui & Sun, 2006). They have the lowest variance in training accuracy, as shown in Table 6. Therefore, the SB-MCF model could integrate the different training advantages of each single classifier.

As shown in Table 8, the testing accuracy and discrete degree of the SB-MCF model is significantly superior to those of DT at the 1% level. The training accuracy of the SB-MCF model is significantly superior to that of all the single classifiers at the 5% level at least, except that of Logit. The SB-MCF model is significantly superior to DT in terms of testing accuracy and discrete degree at the 10% level at least. Thus, from a statistical perspective, the SB-MCF model is significantly superior to each single classifier in terms of testing accuracy and discrete degree.

Additionally, we also analyze the mean testing accuracy class by class. The results are presented in Table 9. Furthermore, we performed *t*-test for the average testing accuracy of each risk level between SB-MCF model and other single classifiers. And the data show that the testing accuracy of SB-MCF model is significantly superior to that of all the single classifiers. In order to avoid redundancy, we do not report the result of *t*-test in this paper.

The experimental results show that the SB-MCF model can improve the average prediction accuracy and reduce the discrete degree of prediction accuracy on different datasets. This model inherited the high recognition performance of Logit and CBR and the stability of SVM; additionally, it avoided the low recognition performance of DT and MDA and the instability of all the basic classifiers (except SVM). Therefore, the SB-MCF method could improve the product-harm risk prediction accuracy by using the complementary information of the different basic classifiers, taking full advantage of the benefits of the basic classifiers and avoiding their disadvantages.

Here insert Tables 7, 8 and 9

5. 2. Comparison of prediction performance between SB-MCF model and other multi-classifier fusion methods

In order to test the performance of the SB-MCF model, we further compared the prediction performance of this model with that of different widely used multi-classifier fusion methods, such as random forests (RF) (Breiman, 2001), majority voting (MAJ) (Ruta & Gabrys, 2005), Bayesian method (Kurzynski & Wozniak, 2012; Woźniak et al., 2014), genetic algorithm (GA) (Kim, Kim & Lee, 2003), Borda counting method (Ho, Hull & Srihari, 1994), and least squares (LS) (Ting & Witten, 1999). The detailed algorithms of the six classifiers fusion methods above are shown as Appendix B. The method for the division of the sample into the training set and testing set was the same as the method used earlier. In this paper, the basic classifier pool size was set to six. We conducted twenty fusion experiments on the twenty experimental datasets using the fusion methods above respectively, and counted the number of basic classifiers selected by the SB-MCF model and the other six fusion methods during each fusion experiment, as shown in Figure 4. The results show that the other six fusion methods (such as RF, MAJ, Bayesian, GA, Borda counting, LS) will fuse all the classifiers in the basic classifier pool during each fusion. However, unlike the other six fusion methods, the SB-MCF method always select some of the most appropriate basic classifiers from a given basic classifier pool adaptively. And the number of basic classifiers selected in every experiment was stable. For example, among twenty fusion experiments, four basic classifiers were selected in twelve fusion experiments, and five basic classifiers were selected in eight fusion experiments. This fully demonstrates that the SB-MCF algorithm is characterised by the adaptive selection of basic classifiers.

Here insert Figure 4

The average prediction accuracy, variances, and discrete coefficients of the different fusion methods on the 20 experimental datasets are shown in Table 10. Compared to the other fusion methods, the SB-MCF model has the highest average testing accuracy and the lowest variances and discrete coefficients on the 20 experimental datasets. Consistent with the results of Table 6, the training and forecasting performances of each model that adopted the nine indicators after reduction as inputs are superior to the performances of those that adopted the 17 initial indicators as inputs. Additionally, we analysed the differences in the accuracy and discrete degree of the different fusion methods on the experimental datasets using the *t*-test; the results are shown in Table 11. The data shows that the SB-MCF model is significantly superior to the other fusion methods in terms of testing accuracy and discrete degree. Additionally, we also analyze the mean testing accuracy class by class. The results are presented in Table 12. Similarly, we also performed *t*-test for the mean testing accuracy of each risk level between SB-MCF model and other fusion methods. The results show that the testing accuracy of SB-MCF model is significantly superior to that of all other fusion methods. In order to avoid redundancy, we do not report the result of *t*-test in this paper. In summary, the performance of the SODM-based multiple classifiers fusion model in

relation to product-harm risk prediction is superior to that of other fusion methods.

Comparing Table 10 and Table 6, the multiple classifiers fusion model did not necessarily outperform the best single classifiers. In this study, the average prediction accuracies of GA and LS were found to be slightly lower than those of single classifiers such as Logit, NN, and CBR. Further, the standard deviation and coefficients of variation of some of the multiple classifiers models were greater than those of some of the single classifiers. We found that this conclusion, an ensemble composed of all the basic classifiers does not necessarily improve classification accuracy, was consistent with the finding reported by Geng et al. (2015); however, it contradicted the result reported by Sun and Li (2008a) and Xiao et al. (2012). According to Table 8 and Table 11, the prediction performance of the SB-MCF model are superior to both the single classifiers and other multiple classifiers fusion model. This further verified the effectiveness of the SB-MCF model.

Here insert Tables 10, 11 and 12

6. Conclusion

The empirical results indicate that the SODM-based multiple classifiers fusion model has higher prediction accuracy and prediction stability compared to those of widely used single classifiers (e.g. logistic regression, neural networks, support vector machines) and other fusion methods (e.g. majority voting, Bayesian method, genetic algorithm). Further, the proposed model has good adaptive fusion features, and it can adaptively select an appropriate subset from the pool of base classifiers for the fusion, which can improve the prediction accuracy of the model. Thus, the proposed model is effective and practical for risk forecasting related to product-harm events.

Reducing the condition attributes using rough set theory, we obtained nine valid risk-forecasting indexes without the loss of information. These indexes are related to the sensitivity of the event, the credibility of the media, enterprise scale, the risk awareness of the enterprise, the degree of public involvement, the frequency of media reports, the relationship between the enterprise and the government, the relationship with suppliers and distributors, and working capital. Additionally, we divided the risk level of a product-harm event into giant warning, heavy warning, medium warning, and light warning levels using fuzzy cluster analysis, which provides an important basis for enterprises and the government to define the risk level of a product-harm event.

In this study, the clustering analysis of product-harm risk was based on a sample of listed companies, and the analysis was limited to the use of stock market performance indicators. The role of financial and non-financial indicators, such as main business revenue fluctuation and customer churn rate, was not considered in this study. Future research could use financial and non-financial indicators in the clustering analysis to improve the risk grading results. Additionally, unlisted companies could be included in the samples of an empirical study on risk grading and early warning systems.

Acknowledgments

This research is supported by the Fundamental Research Funds for the Central Universities (2013RC21), the Natural Science Foundation of China (71573252), and the National Science-technology Support Plan Project (2015BAK10B01). I would also like to thank editors and reviewers for their extremely useful and encouraging comments.

References

AMIRIAN, E., J. Y. LEUNG, S. ZANON and P. DZURMAN (2015) Integrated cluster analysis and artificial neural network modeling for steam-assisted gravity drainage performance prediction in heterogeneous reservoirs, *Expert Systems with Applications*, 42, 723-740.

AVENT, T. and D. LAUFER (2014) Exploring the effectiveness of creating regulatory fit in crisis communications: Can it change perceptions of media coverage during a crisis? *Business Horizons*, 58, 149-156.

BASHEER, I. A. and M. HAJMEER (2000) Artificial neural networks: fundamentals, computing, design, and application, *Journal of microbiological methods*, 43, 3-31.

BONARDI, J. P. and G. D. KEIM (2005) Corporate political strategies for widely salient issues, *Academy of Management Review*, 30, 555-576.

BOREIKO, D. (2003) EMU and accession countries: Fuzzy cluster analysis of membership, *International Journal of Finance & Economics*, 8, 309-325.

BREIMAN, I. (2001) Random Forests, *Machine Learning*, 45, 5-32.

BROWN, S. J. and J. B. WARNER (1985) Using daily stock returns, *Journal of financial economics*, 14, 3-31.

CARVALHO, S. W., E. MURALIDHARAN and H. BAPUJI (2014) Corporate social 'Irresponsibility': Are consumers' biases in attribution of blame helping companies in product–harm crises involving hybrid products? *Journal of Business Ethics*, 1-13.

CHEN, J., H. HONG, W. JIANG and J. D. KUBIK (2013) Outsourcing mutual fund management: firm boundaries, incentives, and performance, *The Journal of Finance*, 68, 523-558.

CLAEYS, A. S. and V. CAUBERGHE (2015) The role of a favorable pre-crisis reputation in protecting organizations during crises, *Public Relations Review*, 41, 64-71.

CLEEREN, K. (2014) Using advertising and price to mitigate losses in a product-harm crisis, *Business Horizons*, 58, 157-162.

COSTEA, A. (2013) Performance benchmarking of non-banking financial institutions by means of self-organising

map-algorithm, *Journal of Economics and Business*, 16, 37-58.

DAWAR, N. and M. M. PILLUTLA (2000) Impact of product-harm crises on brand equity: The moderating role of consumer expectations, *Journal of Marketing Research*, 37, 215-226.

De Mántaras, R. L. (1991) A distance-based attribute selection measure for decision tree induction, *Machine learning*, 6, 81-92.

EASON, K. (2014) Afterword: the past, present and future of sociotechnical systems theory, *Applied ergonomics*, *45*, 213-220.

EINWILLER, S. A., A. FEDORIKHIN, A. R. JOHNSON and M. A. KAMINS (2006) Enough is enough! When identification no longer prevents negative corporate associations, *Journal of the Academy of Marketing Science*, 34, 185-194.

EL-MELEGY, M. T. and S. M. AHMED (2007) Neural networks in multiple classifier systems for remote-sensing image classification, *Springer Berlin Heidelberg*, 65-94.

FENG, H. M. and X. Z. WANG (2015) Performance improvement of classifier fusion for batch samples based on upper integral, *Neural Networks*, 63, 87-93.

GENG, R., I. BOSE and X. CHEN (2015) Prediction of financial distress: An empirical study of listed Chinese companies using data mining, *European Journal of Operational Research*, 241, 236-247.

GOODWIN, R. and S. SUN (2013) Public perceptions and reactions to H7N9 in Mainland China, *Journal of Infection*, 67, 458-462.

HARRIS, R. S., T. JENKINSON and S. N. KAPLAN (2014) Private equity performance: What do we know? *The Journal of Finance*, 69, 1851-1882.

HO, T. K., J. J. HULL and S. N. Srihari (1994) Decision combination in multiple classifier systems, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16, 66-75.

HOVLAND, C. I. and W. WEISS (1951) The Influence of Source Credibility on Communication Effectiveness, *The Public Opinion Quarterly*, 15, 635-650.

HU, H. and M. SATHYE (2015) Predicting financial distress in the Hong Kong growth enterprises market from the perspective of financial sustainability, *Sustainability*, 7, 1186-1200.

HUI, X. F. and J. SUN (2006) An application of support vector machine to companies' financial distress prediction, *Springer Berlin Heidelberg*, 274-282.

IVAKHNENKO, A. G. (1970) Heuristic self-organization in problems of engineering cybernetics, *Automatica*, 6, 207-219.

KIM, E., W. KIM and Y. LEE (2003) Combination of multiple classifiers for the customer's purchase behavior

prediction, *Decision Support Systems*, 34, 167-175.

KIM, M. J. and D. K. KANG (2012) Classifiers selection in ensembles using genetic algorithms for bankruptcy prediction, *Expert Systems with Applications*, 39, 9308-9314.

KLEIN, J. and N. DAWAR (2004) Corporate social responsibility and consumers' attributions and brand evaluations in a product–harm crisis, *International Journal of research in Marketing*, 21, 203-217.

KUMAR, P. R. and V. RAVI (2007) Bankruptcy prediction in banks and firms via statistical and intelligent techniques–A review, *European Journal of Operational Research*, 180, 1-28.

KUNCHEVA, L. I. (2005) Diversity in multiple classifier systems, *Information fusion*, 6, 3-4.

KURZYNSKI, M. and M. WOZNIAK (2012) Combining classifiers under probabilistic models: experimental comparative analysis of methods, *Expert Systems*, 29, 374-393.

LENNOX, C. (1999) Identifying failing companies: a re-evaluation of the logit, probit and MDA approaches, *Journal of Economics and Business*, 51, 347-364.

LIN, J.S. (2012) A novel design of wafer yield model for semiconductor using a GMDH polynomial and principal component analysis, *Expert Systems with Applications*, 39, 6665-6671.

LIU, Y. and V. SHANKAR (2015) The dynamic impact of product-harm crises on brand preference and advertising effectiveness, *Management Science*, 61, 2514-2535.

LIU, C. L. and N. ZHANG (2012) Study on the effectiveness of public company's rumor denial announcements: Evidence from Chinese stock market, *Journal of Management Sciences in China*, 15, 42-54.

LIU, P., S. D. SMITH and A. A. SYED (1990) Stock price reactions to the Wall Street Journal's securities recommendations, *Journal of financial and Quantitative Analysis*, 25, 399-410.

MA, B., L. ZHANG, G. WANG and F. LI (2014) The impact of a product-harm crisis on customer perceived value, *International Journal of Market Research*, 56, 341-366.

MEINTJES, C. and A. F. GROBLER (2014) Do public relations professionals understand corporate governance issues well enough to advise companies on stakeholder relationship management? *Public Relations Review*, 40, 161-170.

MENG, X., T. XING, Y. XING and H. WANG (2014) Application of fuzzy cluster analysis on diagnosing the locations of the hole defects in Acer mono wood using acoustic testing// Software Engineering and Service Science, 5th IEEE International Conference on. IEEE, 958-962.

MEYER, P. A. and H. W. PIFER (1970) Prediction of bank failures, *Journal of Finance*, 25, 853–868.

PARK, B. I., A. CHIDLOW and J. CHOI (2014) Corporate social responsibility: Stakeholders influence on MNEs' activities, *International Business Review*, 23, 966-980.

PARK, J. E. and A. SOHN (2013) The influence of media communication on risk perception and behavior related to mad cow disease in South Korea, *Osong Public Health and Research Perspectives*, 4, 203-208.

PAWLAK, Z. and A. SKOWRON (2007) Rudiments of rough sets, *Information sciences*, 177, 3-27.

RAVISANKAR, P. and V. RAVI (2010) Financial distress prediction in banks using group method of data handling neural network, counter propagation neural network and fuzzy ARTMAP, *Knowledge-Based Systems*, 23, 823-831.

RHEE, M. and P. R. HAUNSCHILD (2006) The liability of good reputation: A study of product recalls in the US automobile industry, *Organization Science*, 17, 101-117.

RUTA, D. and B. GABRYS (2005) Classifier selection for majority voting, *Information fusion*, 6, 63-81.

SAHA, S., C. A. MURTHY and S. K. PAL (2007) Rough set based ensemble classifier for web page classification, *Fundamental Information*, 76, 171-187.

SHEIKHOLESLAMI, M., F. B. SHEYKHOLESLAMI, S. KHOSHHAL, H. MOLA-ABASIA, D. D. GANJI and H. B. ROKNI (2014) Effect of magnetic field on Cu–water nanofluid heat transfer using GMDH-type neural network, *Neural Computing and Applications*, 25, 171-178.

SO, Y. and W. F. KUHFELD (1995) Multinomial logit models, *In SUGI 20 Conference Proceedings*, 1227-1234.

SUN, J. and H. LI (2008a) Data mining method for listed companies' financial distress prediction, *Knowledge-Based Systems*, 21, 1-5.

SUN, J. and H. LI (2008b) Listed companies' financial distress prediction based on weighted majority voting combination of multiple classifiers, *Expert Systems with Applications*, 35, 818-827.

SUN, J., H. LI, P. C. CHANG and K. Y. HE (2015) The dynamic financial distress prediction method of EBW-VSTW-SVM, *Enterprise Information Systems*, 1-28.

SWEENY, K. (2008) Crisis decision theory, *Psychological Bulletin*, 134, 61-76.

TING, K. M. and I. H. WITTEN (1999) Issues in stacked generalization, *J. Artif. Intell. Res.*, 10, 271-289.

TSENG, H. E., C. C. Chang and S. H. Chang (2005) Applying case-based reasoning for product configuration in mass customization environments, *Expert Systems with Applications*, 29, 913-925.

TSAI, C. F. (2008) Financial decision support using neural networks and support vector machines, *Expert Systems*, 25, 380-393.

TSAI, C. F. and Y. F. HSU (2013) A Meta-learning Framework for Bankruptcy Prediction, *Journal of Forecasting*, 32, 167-179.

VASSILIKOPOULOU, A., G. SIOMKOS, K. CHATZIPANAGIOTOU and A. PANTOUVAKIS (2009) Product-harm crisis management: Time heals all wounds? *Journal of Retailing and Consumer Services*, 16,

174-180.

WALUMBWA, F. O., M. A. MAIDIQUE and C. ATAMANIK (2014) Decision-making in a crisis: What every leader needs to know, *Organizational Dynamics*, 43, 284-293.

WANG, D. L., X. F. SONG, J. Y. YUAN and W. Y. YIN (2015) Forecasting Core Business Transformation Risk Using the Optimal Rough Set and the Neural Network. *Journal of Forecasting*, 34, 478-491.

WANG, J. and B. W. RITCHIE (2013) Attitudes and perceptions of crisis planning among accommodation managers: Results from an Australian study, *Safety science*, 52, 81-91.

WESTON, J. and C. WATKINS (1998) Multi-class support vector machines. *Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London*, May.

WHELAN, J. and N. DAWAR (2014) Attributions of blame following a product-harm crisis depend on consumers' attachment styles, *Marketing Letters*, 25, 1-10.

WOZNIAK, M., M. GRANA and E. CORCHADO (2014) A survey of multiple classifier systems as hybrid systems, *Information Fusion*, 16, 3-17.

XIAO, Z., X. YANG, Y. PANG and X. DANG (2012) The prediction for listed companies' financial distress by using multiple prediction methods with rough set and Dempster–Shafer evidence theory, *Knowledge-Based Systems*, 26, 196-206.

YANG, L. Y. (2011) Classifiers selection for ensemble learning based on accuracy and diversity, *Procedia Engineering*, 15, 4266–4270.

YUE, L., W. ZUO, T. PENG, Y. WANG and X. HAN (2015) A fuzzy document clustering approach based on domain-specified ontology, *Data & Knowledge Engineering*, 100, 148-166.

YULE, G. U. (1900) On the association of attributes in statistics: with illustrations from the material of the childhood society, & c. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 1 94, 257-319.

ZHAO, D., F. WANG, J. WEI and L. LIANG (2013) Public reaction to information release for crisis discourse by organization: Integration of online comments, *International Journal of Information Management*, 33, 485-495.

ZHAO, L., Q. WANG, J. CHENG, D. ZHANG, T. MA, Y. C. CHEN and J. J. WANG (2012) The impact of authorities' media and rumor dissemination on the evolution of emergency, *Physica A: Statistical Mechanics and its Applications*, 391, 3978-3987.

ZHANG, M., L. D. XU, W. X. ZHANG and H. Z. LI (2003) A rough set approach to knowledge reduction based on inclusion degree and evidence reasoning theory, *Expert Systems*, 20, 298-304.

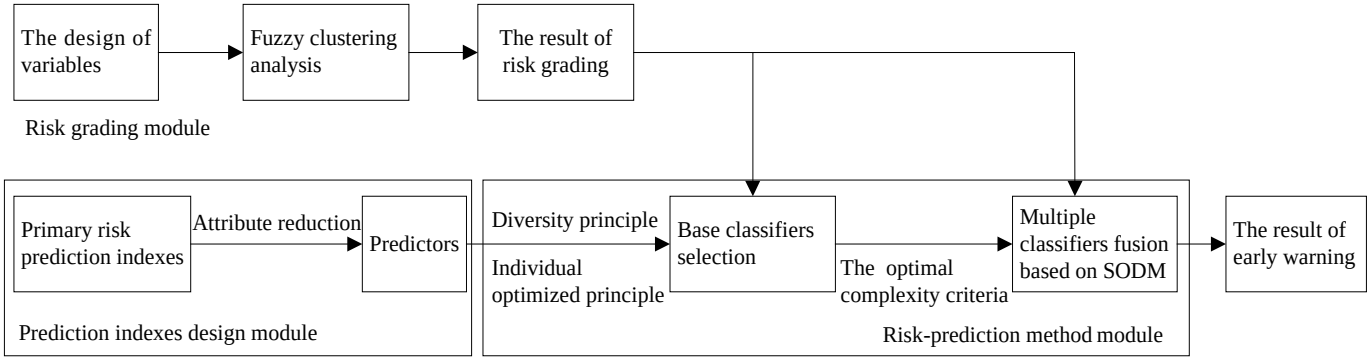

Fig.1. Proposed framework for forecasting the risk of product-harm events

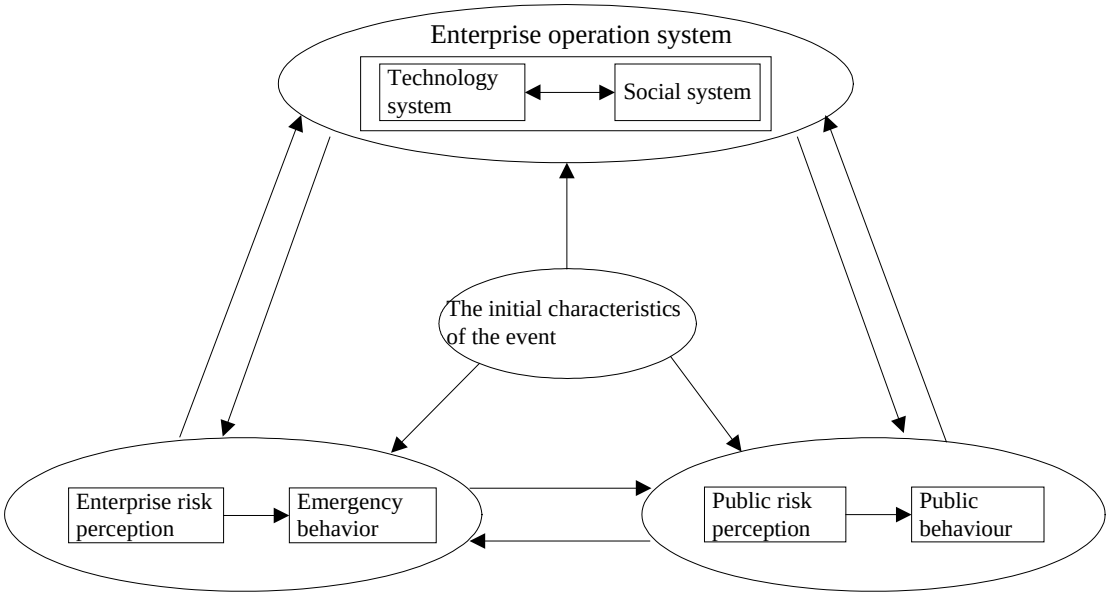

Fig.2. Drivers of the evolution of a product-harm event

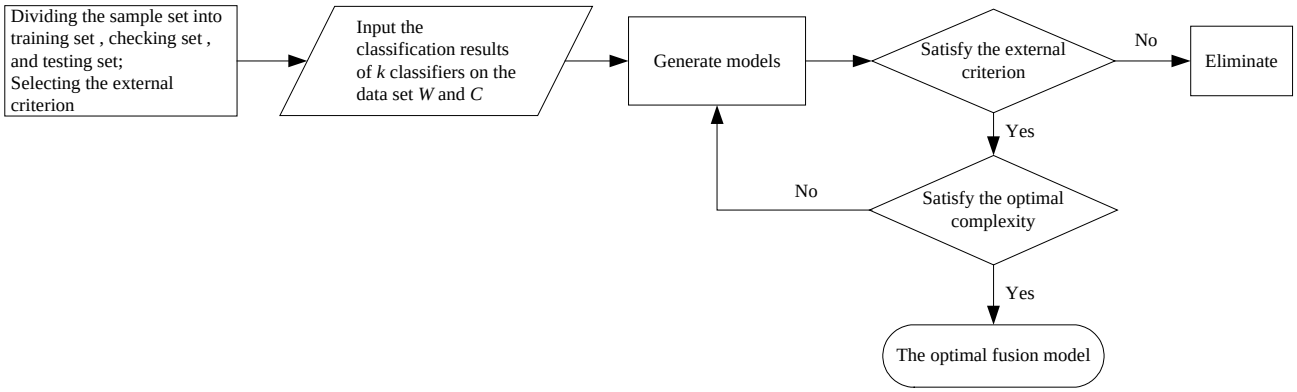

Fig.3. The basic flow of SB-MCF algorithm.

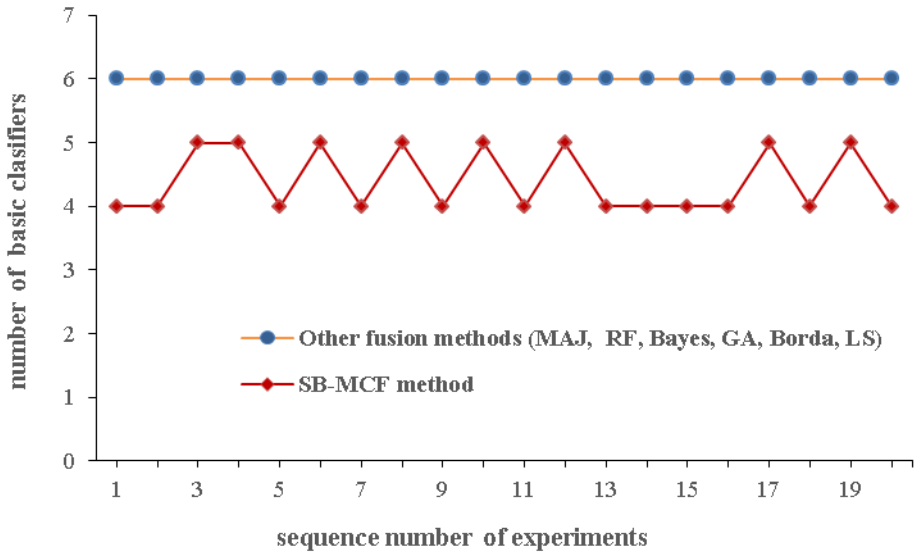

Fig.4. Selection of basic classifiers using different fusion algorithms

Table 1 Initial index system for risk prediction of a product-harm event

Risk factors	Indicators	Definitions and measures
The initial characteristics of the event	The sensitivity of the event (I_1)	The sensitivity of the public to the product quality problem that the event involved
	The influence of media (I_2)	The credibility of the media and the reliability of the information
	Degree of loss (I_3)	The direct loss caused by the event
	Enterprise scale (I_4)	Napierian logarithm of total assets
The risk of enterprise emergency decision	The level of emergency preparedness (I_5)	The perfective of measure or action plan that was constituted in advance by the enterprise to cope with a product-harm crisis
	Risk awareness (I_6)	The crisis awareness of an enterprise to an emergency under daily operating conditions
	Pressure level (I_7)	The attention that the same industry paid to such an event under daily operating conditions
	Response speed (I_8)	The interval between the first report of the event to the first response of the enterprise
The risk of public aggressive behaviour	Customer loyalty (I_9)	The degree of customers' repeat purchases because of their brand preference under normal conditions
	Degree of public involvement (I_{10})	The degree of public participation in commenting on and spreading information about the event
	Frequency of media reports (I_{11})	The frequency of media reports about the product-harm event
The adaptability of enterprise operation system	Political relationship (I_{12})	The compactness degree of the relationship with the government or government officials
	Relationship with suppliers and distributors (I_{13})	The harmonious degree of the relationship with suppliers and distributors
	Relationship with employees (I_{14})	The harmonious degree of the relationship with employees
	Working capital (I_{15})	The capital to be invested in floating assets
	Quick ratio (I_{16})	The ability of an enterprise to reimburse current liabilities
	Interest cover ratio (I_{17})	The ability of an enterprise to repay the loan interest

Table 2 Sample selection process

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	total
The companies collected by the investigation center of Sina.com	26	28	35	43	49	55	62	69	78	30	475
The companies added from the investigation center of Sohu.com and NetEase.com	5	6	3	4	6	5	4	3	5	2	43
Eliminating the companies which are not listed in Shanghai and Shenzhen securities exchange	12	14	11	23	27	20	25	33	38	12	215
Eliminating the companies whose exposure time of product-harm is not clear	5	7	3	6	8	5	7	4	9	3	57
Eliminating the companies whose data is not complete	3	5	7	5	9	11	13	10	15	3	81
The final samples	11	8	17	13	11	24	21	25	21	14	165

Table 3 Descriptive statistics and variance analysis of the variables after classification

Risk levels	Statistics	CAR_{am}	CAR_{sm}	CAR_{ss}	CAR_{sr}
Giant warning	N	38	38	38	38
	Mean	-13.314	-13.886	-14.129	-13.350
	Median	-13.692	-14.850	-14.617	-13.480
	Std. Deviation	1.137	1.698	1.371	1.379
	Minimum	-15.219	-15.887	-16.198	-15.656
	Maximum	-12.043	-11.928	-12.582	-12.921
Heavy warning	N	47	47	47	47
	Mean	-9.352	-9.137	-9.275	-9.819
	Median	-9.391	-8.998	-9.506	-10.284
	Std. Deviation	1.741	1.673	1.431	1.185
	Minimum	-11.369	-11.766	-11.495	-11.078
	Maximum	-6.26	-6.19	-6.69	-7.68
Medium warning	N	45	45	45	45
	Mean	-4.176	-4.375	-4.333	-4.396
	Median	-4.290	-4.371	-4.616	-4.422
	Std. Deviation	-1.086	-0.770	-0.953	-1.350
	Minimum	-5.475	-5.381	-5.731	-6.528
	Maximum	-2.219	-3.108	-2.725	-2.507
Light warning	N	35	35	35	35
	Mean	-0.282	-0.256	-0.583	-0.532
	Median	-0.342	-0.335	-0.623	-0.569
	Std. Deviation	1.907	1.563	1.517	1.662
	Minimum	-3.426	-2.643	-2.644	-2.659
	Maximum	1.652	1.997	1.454	1.063
	F	94.828	121.482	134.315	127.074
	P	0.000	0.000	0.000	0.000

Table 4 Quantitative method of indicators and discretisation intervals

Types	Symbols	Quantitative methods	Discretization Intervals			
			1	2	3	4
The initial characteristics of the events	I_1	The search index of Baidu index(ten thousand)	[0,0.95)	[0.95,1.88)	[1.88,2.86)	[2.86, +∞)
	I_2	The media influence index	[0,55)	[55,80)	[80,100]	——
	I_3	The percentage of respondents who think the reported product quality problems is serious (%)	[0,0.30)	[0.30,0.45)	[0.45,0.80)	[0.80,1]
	I_4	Napierian logarithm of total assets (hundred million)	[0,15.70)	[15.70,25.25)	[25.25,30.5)	[30.55,+∞)
The risk of enterprise emergency decision	I_5	Using word frequency analysis method to count the number of occurrences of the emergency plan, emergency measures and emergency management in company annual reports a year before the event (time)	[0,2)	[2,5)	[5,+∞)	——
	I_6	Using word frequency analysis method to count the number of occurrences of quality, customer, risk and crisis in company annual reports a year before the event (time)	[0,3)	[3,5)	[5, +∞)	——
	I_7	Using word frequency analysis method to count the average number	[0,4)	[4,7)	[7, +∞)	——

		of occurrences of product quality, risk and crisis in company annual reports of the first four companies in the same industry a year before the event (time)				
	I_8	The days from the event was first reported to the companies' first response (day)	[0, 1)	[1,3)	[3,7]	[7,+∞)
The risk of public aggressive behavior	I_9	The percentage of respondents who said no longer buy the products of involved enterprise (%)	[0,0.35)	[0.35,0.52)	[0.52,0.85)	[0.85,1]
	I_{10}	The number of people participated in questionnaire (ten thousand)	[0,2.50)	[2.50,3.80)	[3.80,5.50)	[5.50, +∞)
	I_{11}	Media index of Baidu index	[0,260)	[260,510)	[510,780)	[780, +∞)
The adaptability of enterprise operation system	I_{12}	The proportion of board numbers with political connections (%)	[0, 0.15)	[0.15, 0.33)	[0.33,1]	——
	I_{13}	Represented by supplier, consumer rights responsibility index of Hexun.com	[0, 57)	[57, 82)	[82,100]	——
	I_{14}	Employees responsibility index	[0, 35)	[35,60)	[60,80)	[80,100]
	I_{15}	The gap between the current assets and current liabilities (hundred million)	(-∞, 1.32)	[1.32, 2.50)	[2.50,4.50)	[4.50,+∞)
	I_{16}	The ratio of quick assets dividing by current liabilities (%)	[0, 0.33)	[0.33,1.05)	[1.05, +∞)	——
	I_{17}	The ratio of corporate earnings before interest and tax dividing by interest expense (%)	(-∞, -0.53)	[-0.53,2.80)	[2.80, 6.55)	[6.55,+∞)

Notes: All financial indicators data reflect values from the previous quarter before the events were reported for the first time. The data from the investigation centre of Sina.com and Baidu index reflect the values from the day that the event was reported. All the other indicators data report values from the preceding year before the events.

Table 5 *Q*-statistics results

Classifiers	Logit	NN	DT	SVM	CBR
MDA	0.951	0.943	0.899	0.965	0.957
Logit	—	0.960	0.904	0.963	0.930
NN	—	—	0.918	0.952	0.955
DT	—	—	—	0.927	0.946
SVM	—	—	—	—	0.950

Table 6 Mean accuracy, variance, and discrete coefficients of SB-MCF model and single classifiers

Predictive indexes		Models	MDA	Logit	NN	DT	SVM	CBR	SB-MCF
I_1、I_2、I_4、I_6、I_9、	The results of	Mean	85.618	85.432	89.186	85.533	88.269	—	89.590
I_{11}、I_{12}、I_{13}、I_{15}	training sample	Std. Deviation	1.679	3.358	5.157	7.862	1.525	—	1.176
		Discrete coefficient	0.019	0.039	0.058	0.092	0.017	—	0.013
	The results of	Mean	80.681	84.290	82.118	80.552	82.367	83.360	85.966
	testing sample	Std. Deviation	8.883	8.915	6.794	11.256	5.203	7.856	3.168
		Discrete coefficient	0.110	0.106	0.083	0.140	0.063	0.094	0.037
17 initial indexes	The results of	Mean	84.973	84.726	87.382	84.112	86.575	—	88.029
	training sample	Std. Deviation	2.935	3.803	5.947	8.197	2.696	—	1.810
		Discrete coefficient	0.035	0.045	0.068	0.097	0.031	—	0.021
	The results of	Mean	80.136	83.856	81.891	80.213	81.799	83.182	84.863
	testing sample	Std. Deviation	8.912	9.238	7.159	11.923	5.869	8.403	4.750
		Discrete coefficient	0.111	0.110	0.087	0.149	0.072	0.101	0.056

Table 7 *t*-test results of global training accuracy and discrete degree of each model (*p*-value)

Models	Indexes	Logit	NN	DT	SVM	SB-MCF
MDA	Accuracy	0.586	0.000***	0.704	0.004***	0.000***
	Discrete degree	0.052*	0.009***	0.000***	0.905	0.108
Logit	Accuracy	—	0.000***	0.899	0.002***	0.000***
	Discrete degree	—	0.138	0.001***	0.026**	0.004***
NN	Accuracy	—	—	0.007***	0.087*	0.899
	Discrete degree	—	—	0.033**	0.002***	0.003***
DT	Accuracy	—	—	—	0.003***	0.000***
	Discrete degree	—	—	—	0.000***	0.000***
SVM	Accuracy	—	—	—	—	0.001***
	Discrete degree	—	—	—	—	0.111

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 8 *t*-test results of global testing accuracy and discrete degree of each model (*p*-value)

Models	Indexes	Logit	NN	DT	SVM	CBR	SB-MCF
MDA	Accuracy	0.000***	0.061*	0.736	0.052*	0.038**	0.000***
	Discrete degree	0.885	0.334	0.082*	0.091*	0.416	0.041**
Logit	Accuracy	—	0.043**	0.000***	0.045**	0.236	0.116
	Discrete degree	—	0.337	0.064*	0.098*	0.528	0.043**
NN	Accuracy	—	—	0.318	0.934	0.461	0.008***
	Discrete degree	—	—	0.049**	0.125	0.538	0.086*
DT	Accuracy	—	—	—	0.043**	0.027**	0.000***
	Discrete degree	—	—	—	0.012**	0.053**	0.009***
SVM	Accuracy	—	—	—	—	0.597	0.017**
	Discrete degree	—	—	—	—	0.335	0.121
CBR	Accuracy	—	—	—	—	—	0.047**
	Discrete degree	—	—	—	—	—	0.079*

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 9 Testing accuracy class by class of each model

Indexes	MDA	Logit	NN	DT	SVM	CBR	SB-MCF
Accuracy of light warning class	82.034	85.35	83.257	81.986	83.399	84.690	87.357
Accuracy of medium warning class	79.630	83.101	81.006	79.456	81.427	82.447	84.833
Accuracy of heavy warning class	79.361	83.500	81.249	79.22	81.22	82.115	84.799
Accuracy of giant warning class	81.703	85.241	82.958	81.573	83.421	84.193	86.992

Table 10 Global testing accuracy, variance, and discrete coefficient of different fusion algorithms

Predictive indexes	Statistics	RF	MAJ	Bayes	GA	Borda	LS	SB-MCF
I_1、I_2、I_4、I_6、I_9、I_{11}、I_{12}、I_{13}、I_{15}	Mean	84.002	83.772	82.831	81.575	82.869	81.567	85.966
	Std. Deviation	5.539	5.396	6.437	5.619	7.277	5.183	4.168
	Discrete coefficient	0.066	0.064	0.078	0.069	0.088	0.063	0.049
17 initial indexes	Mean	81.405	81.972	82.151	80.834	81.518	81.095	84.863
	Std. Deviation	6.990	6.213	6.990	6.052	7.944	5.876	5.125
	Discrete coefficient	0.086	0.076	0.085	0.075	0.097	0.072	0.060

Table 11 *t*-test results of testing accuracy and discrete degree of different fusion algorithms

Models	Indexes	RF	Bayes	GA	Borda	LS	SB-MCF
MAJ	Accuracy	0.278	0.222	0.097*	0.661	0.094*	0.096*
	Discrete degree	0.987	0.165	0.879	0.089*	0.931	0.105
RF	Accuracy	—	0.097*	0.056*	0.098*	0.077*	0.099*
	Discrete degree	—	0.270	0.881	0.092*	0.911	0.087*
Bayes	Accuracy	—	—	0.519	0.986	0.496	0.083*
	Discrete degree	—	—	0.401	0.701	0.091*	0.062*
GA	Accuracy	—	—	—	0.537	0.873	0.011**
	Discrete degree	—	—	—	0.104	0.364	0.099*
Borda	Accuracy	—	—	—	—	0.516	0.090*
	Discrete degree	—	—	—	—	0.083*	0.039**
LS	Accuracy	—	—	—	—	—	0.008***
	Discrete degree	—	—	—	—	—	0.106

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 12 Testing accuracy class by class of different fusion algorithms

Indexes	RF	MAJ	Bayes	GA	Borda	LS	SB-MCF
Accuracy of light warning class	85.210	84.830	84.186	83.234	84.440	83.118	87.357
Accuracy of medium warning class	83.021	82.108	81.396	80.149	81.781	80.494	84.833
Accuracy of heavy warning class	82.512	83.005	81.112	79.999	81.592	79.755	84.799
Accuracy of giant warning class	85.269	85.163	84.632	82.911	83.663	82.886	86.992

Appendix A.

(1) Support Vector Machine (SVM)

SVM is a statistic learning method based on fewer samples, which was initially put forward to solve the two-class classification problem. In order to solve the problem of multi-class pattern recognition, many scholars have researched on multi-class SVM classification algorithms and achieved important results, including 'one-against-rest', 'one-against-one', Decision Directed Acyclic Grach (DDAG), etc. In this paper, we adopt 'one-against-rest' algorithm, details are shown as follows:

Step 1: Let $A=\{(x_1,y_1),\cdots,(x_l,y_l\}\in(X\times Y)^l$ be the training set , C be the testing set, where $x_j\in R^n$, $y_j\in\{1,\cdots,N\}$, $j=1,\cdots,l$, l is the sample size, N is the number of categories (in this paper, $N=4$).

Step 2: Contract N sub-classifiers of SVM, and build the linear optimization model of i^{th} sub-classifier of SVM:

$$\begin{cases}\min\dfrac{1}{2}(w^i)^T w^i+c\sum\limits_{j=1}^{M}\xi_j^i\\ s.t.(\mathrm{w}^i)^T\phi(x_j)+b_i\geq 1-\xi_j^i(y_j=i)\\ (\mathrm{w}^i)^T\phi(x_j)+b_i\leq\xi_j^i-1(y_j\neq i)\\ \xi_j^i\geq 0,\ j=1,2,\cdots,l\end{cases}\tag{1}$$

where c is the penalty factor that can be obtained based on enumeration method (in this paper, $c=100$), ξ_j^i is the relaxation factor that allows x_j to be divided into a wrong class.

Step 3: By solving the linear optimization model, N decision functions can be obtained:

$$f_i(x)=\mathrm{sgn}\left[g_i(x)\right]=\mathrm{sgn}\left[\sum_{sv}\alpha_j y_j K^i(x,x_j)+b_i\right](i=1,2,\cdots,N;\ j=1,2,\cdots,l)\tag{2}$$

where sgn( ) is the Symbolic function, $K^i(x,x_j)$ is the Kernel function used in i^{th} sub-classifier (in this paper, we adopt Radial Basis function), b_i is the threshold of i^{th} category.

Step 4: Input each sample of testing set C to N decision functions, N output values $f_i(x)=\mathrm{sgn}\left[g_i(x)\right], i=1,2,\cdots,N$ can be obtained. If there is only one positive value, take the corresponding category of this value as the category of the input sample. If there is more than one positive value or none positive value, comparing the output value $g_i(x)$, and take the corresponding category of the maximum output value $g_i(x)$ as the category of the input sample.

(2) Logit model (Logit)

Logit model is the earliest discrete choice model and also one of the most widely used model. The traditional Logit model, the binomial Logit model, is used to solve two-classification problem. Considering the risk levels of product-harm event were divided into four levels, we chose multinomial Logit model, and the detailed algorithm is shown as follows:

Step 1: Let A be the training set, C be the testing set.

Step 2: Build the multinomial Logit regression model:

$$\ln\left[\frac{P(\mathrm{Y}=i\mid X=x)}{P(Y=r\mid X=x)}\right]=\alpha_i+x'\beta_i \quad i=1,2,\cdots,N,i\neq r \tag{3}$$

where i stands for different category (in this paper, $N=4$), r stands for the reference category (in this paper, we set the Giant warning as the reference category, that is $r=4$). For the convenience of solving the model, we let the regression coefficient vector β_r of reference category be zero vector. $P(\mathrm{Y}=i\mid \mathrm{X}=x)$ is the posterior probability of the sample x belongs to the known category i.

Step 4: Take regression analysis in training set A, N probability equations can be obtained:

$$P(Y=r\mid X=x)=\frac{1}{1+\sum_{j=1,j\neq r}^{k}e^{\alpha_j+x'\beta_j}}$$
$$P(Y=i\mid X=x)=\frac{e^{\alpha_i+x'\beta_i}}{1+\sum_{j=1,j\neq r}^{k}e^{\alpha_j+x'\beta_j}} \quad i=1,2,\cdots,N,i\neq r \tag{4}$$

where $\beta_1,\cdots,\beta_N$ are the estimates by using maximum likelihood, $\alpha_1,\alpha_2,\cdots,\alpha_N$ are regression intercept.

Step 5: Input each sample of testing set C to Eq. (2), N probability values can be obtained which mean the probability of the sample belongs to the known categories. Comparing the N probability values, and take the corresponding category of maximum probability value as the category of the testing sample.

(3) Multiple Discriminant Analysis (MDA)

MDA is a multivariate statistical method, including maximum likelihood, distance discrimination method, Fisher discrimination method and Bayes discrimination method. In this paper, we adopt distance discrimination method, and the detailed algorithm is shown as follows:

Step 1: Let A be the training set, which include N sample sets $G_1,G_2,\cdots,G_N$ (N is the number of categories, in this paper, $N=4$), C be the testing set. X_i^{α} is the i^{th} sample of G_{α}, where $\alpha=1,2,\cdots,N;i=1,2,\cdots,n$ (n is the sample size).

Step 2: Estimate the mean μ_{α} and covariance matrix E_{α} of G_{α} based on unbiased estimation:

$$\mu_{\alpha}=\bar{X}^{\alpha}=\frac{1}{n_{\alpha}}\sum_{i=1}^{n_{\alpha}}X_i^{\alpha}$$
$$\hat{E}_{\alpha}=\frac{1}{n-1}S_{\alpha}$$
$$S_{\alpha}=\sum_{i=1}^{n_{\alpha}}(X_i^{\alpha}-\bar{X}^{\alpha})(X_i^{\alpha}-\bar{X}^{\alpha})' \quad \alpha=1,2,\cdots,N \tag{5}$$

Step 3: The mahalanobis distance from the sample X to G_{α} can be obtained by Step2:

$$D^2(X,G_{\alpha})=(X-\mu_{\alpha})'\mathrm{E}_{\alpha}^{-1}(X-\mu_{\alpha}),\alpha=1,2,\cdots,N \tag{6}$$

The judging criterion is:

$$X \in G_i \quad if\ D^2(X, G_i) = \min_{1 \le \alpha \le k} D^2(X, G_\alpha) \tag{7}$$

Step 4: Input each sample of the testing set C to Eq. (2) to obtain the mahalanobis distances of the testing sample to each sample sets $G_\alpha (\alpha = 1, 2, \cdots, N)$, and take the corresponding categories of the minimum mahalanobis distance as the category of the testing sample based on the judging criterion.

(4) Neural Network method (NN)

BP neural network is the abbreviation of erroneous reverse transmission neural network, which is presently one of the most widely applied neural network models, the detailed algorithm is shown as follows:

Step 1: Let $A(\mathrm{X}_l, \mathrm{Y}_l)$ be the training set, where l stands for the number of the training sample, C be the testing set.

Step 2: Initialize the weights and thresholds whose values are given randomly, and set $t = 0$.

Step 3: Let $X_l = (x_1^l, x_2^l, \cdots, x_n^l)$ be the given input and $Y_l = (y_1^l, y_2^l, \cdots, y_n^l)$ be the target output, where x_i^l stands for the i^{th} indicator of the l^{th} sample, y_i^l stands for the desired output of the l^{th} sample.

Step 4: Compute the output of three-layers neural network based on sigmoid function. The outputs of the hidden layer and output layer can be calculated respectively from:

$$z_j = \frac{1}{1 + e^{-\sum_{i=0}^{n} w_{ij}^1 x_i}} \quad j = 1, 2, \cdots, m \tag{8}$$

$$y_k = \frac{1}{1 + e^{-\sum_{j=0}^{m} w_{jk}^2 z_j}} \quad k = 1, 2, \cdots, p \tag{9}$$

where n is the nodes in the input layer (in this paper, $n = 10$), m is the nodes in the hidden layer (in this paper, $m = 19$), p is the nodes in the output layer (in this paper, $p = 4$). w_{ij}^1 is the connection weight between input layer and hidden layer. w_{jk}^2 is the connection weight between hidden layer and output layer.

Step 5: Adjust the interconnection weights during training by employing a method known as error back-propagation. The weights change equations on the hidden layer and output layer are respectively:

$$w_{ij}^1(t+1) = w_{ij}^1(t) + \eta^1 \delta_{pj}^1 x_i \tag{10}$$

$$w_{jk}^2(t+1) = w_{jk}^2(t) + \eta^2 \delta_{pk}^2 z_j \tag{11}$$

where η is the learning rate controlling the update step size, δ_{pj}^1 and δ_{pk}^2 are the error terms for hidden units and output units respectively.

If the neuron is in the output layer, δ_{pk}^2 can be calculated from:

$$\delta_{pk}^2 = y_k(1 - y_k)(y_k^0 - y_k) \tag{12}$$

If the neuron is in the hidden layer, δ_{pj}^1 can be calculated from:

$$\delta_{pj}^1(t+1) = z_j(1 - z_j) \sum_{k=0}^{m} \delta_{pk}^2(t) \cdot w_{jk}^2(t) \tag{13}$$

Step 6: Minimize the square error:

$$\Delta E = E(t+1) - E(t) < \varepsilon \tag{14}$$

$$E = \frac{1}{2}\sum_{k=1}^{p}\left[y_k(t) - y_k^0(t)\right]^2 \tag{15}$$

where ε is the maximum allowable error. If $0 \le \varepsilon < 1$ or $t \le T^0$ (T^0 is the cycles), then the training of the network was finished, otherwise, set $t+1 \to t$, and continue to train the neural network.

(5) Decision Tree (DT)

Decision tree is one of well-known classification algorithms, including ID3 algorithm, C4.5 algorithm, CART algorithm, etc. In this paper, we adopt ID3 algorithm, details are shown as follows:

Step 1: Let A be the training set which contains N categories (in this paper, $N = 4$). Based on s values of the attribute B, training set A can be divided into s subsets $A_1, A_2, \cdots, A_s$.

Step 2: Create a root node m. If the samples fall under the same category C_i, let the node be the leaf, and label it with this category C_i, otherwise, turn to Step3.

Step 3: Compute the information gain for each attribute B (that is $gain(B)$), and take the attribute B with the highest information gain as the test attribute of this node:

$$\begin{aligned} gain(\mathrm{B}) &= \inf(\mathrm{A}) - \sum_{i=1}^{s}\frac{|A_i|}{|A|}\inf(\mathrm{A}_i) \\ \inf(\mathrm{A}) &= -\sum_{j=1}^{m} p_j \times \log_2 p_j \end{aligned} \tag{16}$$

where p_j is the probability of category j in training set A.

Step 4: Establish a branch for each test attribute $B = b_i$, and generate a leaf node for the subset A_i.

Step 5: The process above is a recursive procedure until one of the recursion limits is reached:

All samples in the given node fall under the same category;

None attribute left can be used to further divide the sample;

The branch established by attribute b_i has no samples.

(6) Case Based Reasoning (CBR)

The case-based reasoning is a strategy to demonstrate knowledge, and its primary feature is to use previous similar experiences to solve current problems. The detailed algorithm of case-based reasoning is as follows:

Step 1: Let the training set $A = (x_1, x_2, \cdots, x_n)$ be the case set in the case base. $F = \{f_1, f_2, \cdots, f_m\}$ is the index set of the samples (in this paper, $m = 10$). Vector $B_i = \left(b_{i1}, b_{i2}, \cdots b_{im}\right)$ represents the measurement of case x_i for the index set F, where $0 \le b_{ij} \le 1$.

Step 2: Calculate the similarity between the testing sample x^* and the case x_i which has been restored in the case base:

$$sim(x_i, x^*) = \frac{B_i \cdot B^*}{\|B_i\| \cdot \|B^*\|} = \frac{\sum_{j=1}^{m} b_{ij} b_j^*}{\sqrt{\sum_{j=1}^{m} (b_{ij})^2} \sqrt{\sum_{j=1}^{m} (b_j^*)^2}} \tag{17}$$

Step 3: Calculate the differences between the testing sample x^* and the case x_i which has been restored in the case base:

$$dif(x_i, x^*) = \|B_i - B^*\| = \sqrt{\sum_{j=1}^{m} (b_{ij} - b_j^*)^2} \tag{18}$$

Step 4: Calculate the comprehensive similarity between the testing sample x^* and the case x_i:

$$\alpha_i = \frac{sim(x_i, x^*)}{dif(x_i, x^*)} \tag{19}$$

Step 5: Get the maximum comprehensive similarity $\alpha_j = \max(\alpha_i)$, and take the corresponding category of the case x_j as the category of the testing sample.

Appendix B.

(1) Random forests (RF)

Random forests algorithm is a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest, details are shown as follows.

Step 1: The random forest classifier consists of x decision trees, and each tree is composed of y basic classifiers. Each classifier is used to test the risk prediction indexes, and then get $Truth\,(1)$ or $False\,(0)$ (in this paper, $x = 7$, $y = 6$).

Step 2: For each sample in the training set A, each basic classifier determines whether it belongs to giant warning, and then get x $y-bit$ binary digits. These binary digits are subscripts of *x* arrays with y^2 size and these arrays are named G. Every time get such a subscript, the corresponding figure of G adds 1. In the same way, the array H (heaving warning), the array M (medium warning) and the array L (light warning).

Step 3: For a given testing sample, x $y-bit$ binary digits $Ri, i = 1,2,\cdots,x$ can be obtained after it was determined by all the basic classifiers. The posterior probabilities of giant warning, heaving warning, medium warning and light warning can be defined respectively as:

$$PosteriorG[i] = G(Ri)/(G(Ri) + H(Ri) + M(Ri) + L(Ri)) \tag{20}$$

$$PosteriorH[i] = H(Ri)/(G(Ri) + H(Ri) + M(Ri) + L(Ri)) \tag{21}$$

$$PosteriorM[i] = M(Ri)/(G(Ri) + H(Ri) + M(Ri) + L(Ri)) \tag{22}$$

$$PosteriorL[i] = L(Ri)/(G(Ri) + H(Ri) + M(Ri) + L(Ri)) \tag{23}$$

Step 4: Compute the average value of the above four arrays respectively, and take the corresponding class of the largest average value as the final risk grading.

(2) Majority voting (MAJ)

The multi-classifier fusion based on majority voting assuming that each classifier gives a single class label as an output and the final prediction result determined by the most classifiers' outputs, details are shown as follows.

Step 1: Let K be the number of classifiers, N be the number of categories and $\Phi = \{1,2,\cdots,N\}$ be the set of class labels (in this paper, $K = 6$, $N = 4$). For a given sample x, the prediction output by classifier f_k is represented as $f_k(x)$, let binary characteristic function be defined as follows:

$$T_k(x \in c_i) = \begin{cases} 1 & f_k(x) \in i, i \in \Phi \\ 0 & otherwise \end{cases} \tag{24}$$

Step 2: The final risk grading of sample x can be defined as:

$$F(x) = j \quad if \ T_F(x \in c_j) = \max_{i \in \Phi}(T_F(x \in c_i)) \tag{25}$$

(3) Bayesian method

The Bayesian method can be applied to classifiers fusion under the condition that the outputs of classifiers are expressed in posterior probabilities, details are shown as follows.

Step 1: Let K be the number of classifiers, N be the number of categories and $\Phi = \{1,2,\cdots,N\}$ be the set of class labels (in this paper, $K = 6$, $N = 4$). Suppose the confusion matrix of classifier f_k is denoted as CM_k.

$$CM_k = \begin{bmatrix} n_{11}^k & n_{12}^k & \cdots & n_{1N}^k \\ n_{21}^k & \ddots & \cdots & n_{2N}^k \\ \vdots & \vdots & \cdots & \vdots \\ n_{N1}^k & n_{N2}^k & \cdots & n_{NN}^k \end{bmatrix} \tag{26}$$

When $i = j$, n_{ij}^k represents the number of samples which belong to class c_i and are correctly classified as c_i by classifier f_k. When $i \neq j$, n_{ij}^k represents the number of samples which belong to class c_i but are misclassified as c_j by classifier f_k.

Step 2: The number of samples which are classified as c_j by classifier f_k is denoted as n_j^k:

$$n_j^k = \sum_{i=1}^{N} n_{ij}^k, \quad j = 1,2,\cdots,N \tag{27}$$

Step 3: Under the precondition that sample x is classified as class c_j by classifier f_k, the conditional probability of this sample truly belonging to class c_j should be represented as follows:

$$P(x \in c_i \mid f_k(x) = j) = \frac{n_{ij}^k}{n_j^k}, \quad i, j = 1,2,\cdots,N \tag{28}$$

Step 4: The probability of this sample x belonging to class c_j should be represented as follows:

$$P_F(x \in c_i) = Z \prod_{k=1}^{K} P(x \in c_i \mid f_k(x) = j) \quad i, j = 1,2,\cdots,N \tag{29}$$

where $\frac{1}{Z}=\sum_{i=1}^{N}\prod_{k=1}^{K}P(x\in c_i \mid f_k(x)=j)$ is a normalization coefficient that satisfies $\sum_{i=1}^{N}P_F(x\in c_i)=1$.

Step 5: The final risk grading of sample x can be defined as:

$$F(x)=j \quad if \ P_F(x\in c_j)=\max_{i\in\Phi}P_F(x\in c_i) \tag{30}$$

(4) Genetic algorithm (GA)

GA-based multiple classifiers combination method that integrates the measurement level classification results generated by multiple classifiers into a single result, details are shown as follows.

Step 1: Generate initial population o_i. Let K be the number of classifiers, N be the number of categories and $\Phi=\{1,2,\cdots,N\}$ be the set of class labels (in this paper, $K=6$, $N=4$). For a given sample x, each classifier representing it by a measurement vector $M_k=\{m_{1k},m_{2k},\cdots,m_{Nk}\}$, where $k=1,2,\cdots,K$ and m_{ik} is the measurement value of k^{th} classifier for class c_i. Suppose that $W_k=\{w_{1k},w_{2k},\cdots,w_{Nk}\}$ is the weight vector representing the relative significance of k^{th} classifier for all categories, where $k=1,2,\cdots,K$ and w_{ik} is the degree of importance of k^{th} classifier for class c_i.

$$o_i=\sum_{k=1}^{K}w_{ik}m_{ik}, \quad i=1,2,\cdots,N \tag{31}$$

Step 2: The fitness function is defined as follows:

$$Fitness(WS_q)=\left[\frac{\sum_{i=1}^{s}ADD(WS_q)}{S}\right] \tag{32}$$

$$ADD(WS_q)=\begin{cases}1, & if\ correctly\ matched\\ \lambda\left[\dfrac{o_j(WS_q)}{\sum_{i=1}^{N}o_i(WS_q)}\right], & otherwise\end{cases} \tag{33}$$

where S is the total number of training data and $o_i(WS_q)=\sum_{k=1}^{K}w_{ik}m_{ik}$, λ is the constant to control the influence of potential hit on overall learning process and j is the true class for the input (in this paper, $\lambda=0.5$).

Step 3: Selection operator. The probability of selecting candidate solution $P(WS_q)$ is given by

$$P(WS_q)=\frac{Fitness(WS_q)}{\sum_{i=1}^{M}Fitness(WS_i)} \tag{34}$$

where M is the constant denoting the population size (in this paper, $M=150$).

Step 4: Crossover and mutation operators. In order to introduce variation into the new offspring, we apply

the crossover and mutation operators to the individuals of the current population.

Crossover involves the mixing of two individuals to yield two new ones. The mutation operator selects some elements of an individual at random based on the mutation rate and adds a random value to it. This operation ensures the diversity in the weight matrices over long periods of time and prevents stagnation in the convergence of the optimization.

Step 5: Repeat Step3 and Step4 until one of the following two conditions is met:

The number of iterations is greater than the set threshold η (in this paper, $\eta = 100$);

$|\max(Fitness(P(gen))) - Fitness(P(gen-1))| < \delta$, where δ is a pre-set positive decimal (in this paper, $\delta = 0.01$).

Step 6: The final risk grading of sample x can be defined as:

$$F(x) = j, \quad o_j^* = \max_{i \in \Phi} o_i^* \tag{35}$$

(5) Borda counting method

Borda counting for a class is the sum of the number of categories ranked below it by each classifier, details are shown as follows:

Step 1: Let K be the number of classifiers, N be the number of categories and $\Phi = \{1,2,\cdots,N\}$ be the set of class labels (in this paper, $K = 6$, $N = 4$). Rank the classes according to the output vector of classifier f_k, then assign the first class, the second class, …, the last class the value of $(N-1)$ to 0 respectively. For a given sample x, let b_{kj} be the score of k^{th} classifier for j^{th} class.

Step 2: Calculate the score of sample x belongs to j^{th} class

$$b_j = \sum_{k=1}^{K} b_{kj} \tag{36}$$

Step 3: Take the corresponding class of $\max_{i \in \Phi}(b_i)$ as the final risk grading.

(6) Least squares (LS)

LS-based multiple classifiers fusion is a special kind of linear fusion that all the weights of basic classifiers are estimated by LS, details are shown as follows:

Step 1: For a given sample x_i in training set A, suppose that $y(x_i)$ is the actual risk grading vector, $y_k(x_i)$ is the measurement vector of classifier f_k. The prediction error information matrix E can be defined as:

$$E = [(e_{ki})_{K \times A}][(e_{ki})_{K \times A}]^T, \quad k = 1,2,\cdots,K; i = 1,2,\cdots,P \tag{37}$$

where $e_{ki} = y(x_i) - y_k(x_i)$ is the prediction error of k^{th} classifier for sample x_i, P is the sample size.

Step 2: Let $W = (w_1, w_2, \cdots, w_K)^T$ be the weight vector of K classifiers, where w_k is the relative significance of k^{th} classifier for all categories. Suppose $e_i = y(x_i) - \hat{y}(x_i), \quad i = 1,2,\cdots,P$ representing the

prediction error of the combination early warning model for sample x_i , where $\hat{y}(x_i) = w_1 y_1(x_i) + w_2 y_2(x_i) + \cdots + w_K y_K(x_i)$ is the combination early warning model. The sum of squared error of LS-based multiple classifiers fusion model can be defined as:

$$S = \sum_{i=1}^{P} e_i^2 = \sum_{i=1}^{P} (\sum_{k=1}^{K} w_k e_{ki})^2 = W^T E W \tag{38}$$

Step 3: According to the minimum square error criterion, the optimal weight coefficients W_{opt} can be determined by the following model:

$$\begin{cases} \min S = W^T E W \\ s.t. \quad R^T W = 1 \\ R^T = (1,1,\cdots,1)_{1\times K} \end{cases} \tag{39}$$

The optimal weight $W_{opt} = \dfrac{E^{-1}R}{R^T E^{-1} R}$ can be obtained by solving the above model.

Step 4: Let N be the number of categories and $\Phi = \{1,2,\cdots,N\}$ be the set of class labels (in this paper, $N = 4$). For a given testing sample x , each classifier representing it by a measurement vector $M_k = \{m_{1k}, m_{2k}, \cdots, m_{Nk}\}$, where $k = 1,2,\cdots,K$ and m_{ik} is the measurement value of k^{th} classifier for class c_i . The output of the LS-based multiple classifiers fusion model can be defined as:

$$G = [M_1, M_2, \cdots M_K] W_{opt} = [g_1, g_2, \cdots g_N]^T \tag{40}$$

where g_i is the measurement value for class c_i .

Step 5: The final risk grading of sample j can be defined as:

$$F(x) = j \quad if \ g_j = \max_{i\in\Phi}(g_i) \tag{41}$$