

# An evidential reasoning-based decision support system for handling customer complaints in mobile telecommunications

Ying Yang<sup>a,b,\*</sup>, Dong-Ling Xu<sup>b,c</sup>, Jian-Bo Yang<sup>b,c</sup>, Yu-Wang Chen<sup>c</sup>

<sup>a</sup> School of Management, Hefei University of Technology, Hefei 230009, China

<sup>b</sup> Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, Hefei, China

<sup>c</sup> Alliance Manchester Business School, The University of Manchester, Manchester M15 6PB, UK

## ARTICLE INFO

### Article history:

Received 31 January 2018

Received in revised form 14 September 2018

Accepted 17 September 2018

Available online 25 September 2018

### Keywords:

Decision support system

Customer complaint handling

Evidential reasoning rule

Classification

Mobile telecommunications

## ABSTRACT

Handling customer complaints is a decision-making process that inherently involves a classification problem where each complaint should be classified exclusively to one of the complaint categories before a resolution is communicated to customers. Previous studies focus extensively on decision support systems (DSSs) to automate complaint handling, while few addresses the issue of classification imprecision when inaccurate or inconsistent information exists in customer complaint narratives. This research presents a novel DSS for handling customer complaints and develops an evidential reasoning (ER) rule-based classifier as the core component of the system to classify customer complaints with uncertain information. More specifically, textual and numeric features are firstly combined to generate evidence for formulating the relationship between customer complaint features and classification results. The ER rule is then applied to combine multiple pieces of evidence and classify customer complaints into different categories with probabilities. An empirical study is conducted in a telecommunication company. Results show that the proposed ER rule-based classification model provides high performance in comparison with other machine learning algorithms. The developed system offers telecommunication companies an informative and data-driven method for handling customer complaints in a systematic and automatic manner.

© 2018 Elsevier B.V. All rights reserved.

## 1. Introduction

Various data in the forms of texts, voices, graphics and videos are delivered by mobile phones at an ever-increasing speed and in a great number of ways. As the service varieties and technical complexities increase, mobile network operators are facing an increasing amount of complaints. 28.5% of respondents in a survey reported with the experience of having problems with telecommunication services, such as delay and coverage in service, incorrect billing, improper charging or other relevant problems [1]. Handling customer complaints means taking different resolutions for specific complaints. The cause of complaints should be figured out and an optimal resolution be selected properly and quickly. Otherwise, it is very easy to end up with dissatisfied and churned customers. Therefore, mobile network operators are confronted with a challenging problem that how to figure out the causes of customer complaints efficiently and effectively so as to provide appropriate resolutions and improve customer satisfaction.

Handling customer complaints in mobile telecommunications can be formulated as a complex decision-making process involving

classification with uncertain information. Customer complaints may result from multiple causes, or so-called different output classes in the context of classification, e.g., network failure or customer terminal failure. Many researchers have developed decision making support techniques to improve handling efficiency. For example, a valid complaint can be distinguished from invalid ones [2], and complaint emails can be filtered out automatically [3] through implementing DSSs. Web-based DSSs are also used to map customer complaints and to measure customer satisfaction [4]. An interoperable ontology is presented for analyzing complaint-related knowledge and handling complaints intelligently [5]. Optimization techniques such as data envelopment analysis are also combined with data mining techniques to recognize valued customers with complaints [6]. It is evident that DSSs can integrate human experience and machine learning techniques for automating complaint processing. However, classification imprecision commonly exists, especially when the complaints are described with inaccurate and even conflicting information due to customers' limited knowledge about service or product failure.

A set of decision-making models have been proposed on the basis of Dempster–Shafer theory (DST) of evidence, fuzzy logic or rough set theory to improve the accuracy of classification under uncertainty. For example, an evidential reasoning framework is developed to aid decision makers when the available knowledge

\* Corresponding author at: School of Management, Hefei University of Technology, Hefei 230009, China.

E-mail address: [hfyonline@126.com](mailto:hfyonline@126.com) (Y. Yang).

may be imprecise, conflicting and uncertain [7]. A rough set approach is also proposed to predict financial ratios for distressed companies [8]. A model named the evidential signal of price herds based on rough set theory is developed to derive characteristics of price herds and predict future prices [9]. The recently-developed evidential reasoning (ER) rule, established based on the Dempster's rule in the DST theory and the original evidential reasoning algorithm, is also proposed for data classification with uncertainty [10]. Nevertheless, textual information is neglected for decision-making in the research above.

In this study, we propose a methodology for classifying customer complaints under uncertainty and ambiguity. An ER-based decision support system is firstly proposed to handle customer complaints in mobile telecommunications. Text mining techniques are used to extract textual features from the narratives of customer complaints. The evidence for capturing the relationships between features and class memberships under uncertainty are obtained from Bayesian statistics and characterized as a belief distribution. Both uncertain and conflicting information related to customer complaints are considered to generate evidence and combined using the ER rule. The developed decision support system is helpful to facilitate the process of handling customer complaints efficiently and further improve the quality of resolution.

The key contributions of this research are two-fold. On the one hand, we developed an ER rule-based classification model for classifying customer complaints. Evidence is obtained from both textual and numeric features and combined using the ER rule. The proposed classification model can potentially be applied to other relevant domains. On the other hand, we presented an intelligent decision support system to deal with inaccurate and inconsistent information for decision-makers, where both structured and unstructured data can be combined simultaneously in complex decision environments.

The remainder of the paper is organized as follows. Previous research on customer complaint management and the basics of the ER rule are briefly introduced in Section 2. An evidential reasoning-based decision support system is developed for combining multi-source information in Section 3. An empirical study is conducted in Section 4 to demonstrate the applicability and effectiveness of the proposed classification model. Discussions and conclusions are given in the last two sections.

## 2. Related work

### 2.1. Customer complaint management

Customer complaints form a critical source of information for improving a firm's products or services. When firms' products or services cannot meet customers' expectation, customers are very likely to become dissatisfied. According to the theory of customer complaint behavior, most dissatisfied customers usually withdraw their patronage and express negative views on products or services to other people. Only a small percentage of dissatisfied customers make a complaint and believe in firms' ability to resolve their problems fairly [11]. Handling these complaints successfully helps not only identify the defects in products or services, but also maintain customer loyalty.

Customer complaint management is a process of recording and resolving customer complaints. Firms often develop a customer relationship management (CRM) to handle their complaints and create recommendations and solutions [12]. Encouraging, receiving and handling complaints, and sending feedback to customers are key functions embedded in the CRM system. Meanwhile, firms may develop various channels to encourage customers to express their dissatisfaction actively. These complaint channels are often flexible and easy to follow, such as via free helplines, links on

firms' websites and online chat softwares including WeChat, QQ or Facebook.

A complaint handling process begins once a customer complaint is received. Procedures have been developed to specify what actions need to be performed by employees who directly interact with dissatisfied customers. A set of communication rules or behavioral norms are established to help employees avoid possible conflicts. After a complaint is settled, feedback will be sent to inform customers about the activities carried out in handling the complaint, the causes that generated the complaint and the measures to be taken in the future [13].

Generally speaking, there are three types of CRM systems adopted by mobile network operators for customer complaint management, including operational, collaborative and analytical CRM systems [6]. An operational CRM system aims to provide a comprehensive view of customers and delivers customer services to enhance the effectiveness of daily interactions and operations. A collaborative CRM system enables a company to directly interact and communicate with customers. An analytical CRM system is developed to analyze the data generated at the operational CRM level. Although CRM systems have been adopted by companies to manage complaints and improve their interactions with customers, few systems make use of data mining or machine learning techniques to improve the performance of handling customer complaints. Current customer complaint handling methods are mostly limited to use keywords or workflow processes to categorize a complaint to a certain domain-specific class manually or semi-automatically [6].

Therefore, one of the strengths of this study is that text mining techniques are used and an evidential reasoning rule-based classification model is proposed to process customer complaints automatically and to improve handling performance.

### 2.2. The evidential reasoning rule

The ER rule provides a scheme to deal with classification imprecision problems from the perspective of uncertain information fusion. It is established to advance Bayes rule and Dempster rule for evidence combination and information fusion. The evidence is extracted from data to represent the relationship between a feature and class memberships. When multiple pieces of evidence are highly or completely conflicting, the ER rule can still be used to fuse them and generate a classification decision while Bayes rule and Dempster rule cannot be applied [14].

The establishment of the ER rule is based on Dempster-Shafer scheme for evidence aggregation. Basic probabilities can be assigned to both singleton propositions and any of their subsets. A piece of evidence is profiled by a belief distribution defined on the power set of the frame of discernment. The belief distribution is regarded as a natural and flexible form of probability distribution and allows inexact reasoning at whatever level of abstraction [15].

A frame of discernment,  $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$ , is adopted in the ER rule scheme to represent all hypotheses. The hypotheses are mutually exclusive and collectively exhaustive. The power set of  $\Theta$  consists of  $2^N$  subsets, denoted by  $P(\Theta)$ , as follows:

$$P(\Theta) = 2^\Theta = \{\Phi, \{\theta_1\}, \dots, \{\theta_N\}, \{\theta_1, \theta_2\}, \dots, \{\theta_1, \theta_N\}, \dots, \{\theta_1, \dots, \theta_{N-1}\}, \Theta\}$$

A basic probability mass that is assigned exactly to a proposition is defined as  $m(\theta)$ . The basic probability mass assigned exactly to  $P(\Theta)$  is referred to as the degree of global ignorance and that assigned exactly to a smaller subset of  $\Theta$  except for any singleton proposition or  $\Theta$  is called the degree of local ignorance.

A belief distribution function is used to measure the extents to which the evidence supports each hypothesis and the subsets of  $P(\Theta)$ . Such a distribution is also called a piece of evidence

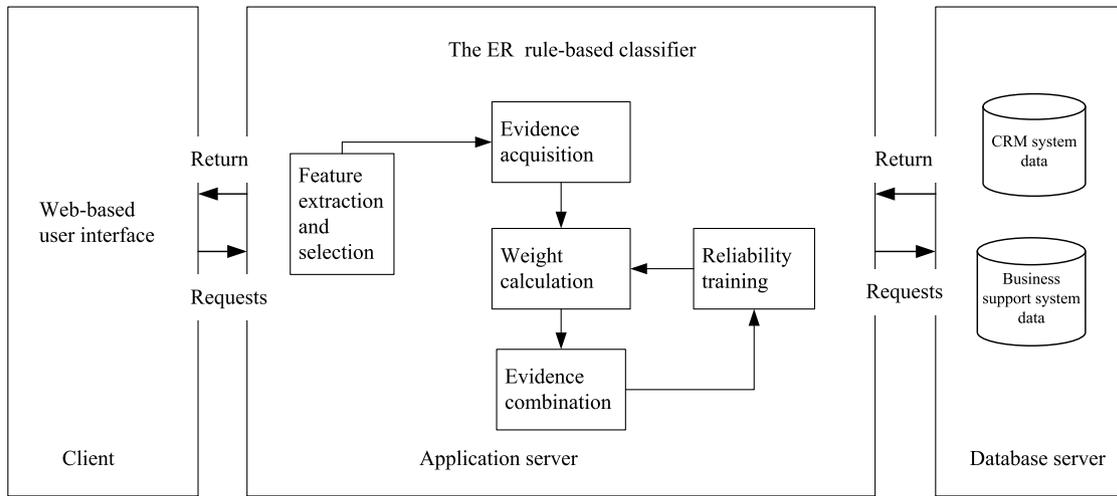


Fig. 1. The architecture of the evidential reasoning-based decision support system.

[14]. Therefore, a piece of evidence  $e_j$  can be profiled by a belief distribution defined on the power set of the frame of discernment  $\Theta$ , as follows:

$$e_j = (\theta, p_{\theta,j}), \forall \theta \subseteq \Theta, \sum_{\theta \subseteq \Theta} p_{\theta,j} = 1$$

where  $\theta$  is a proposition which can be a subset of  $\Theta$  or an element of  $P(\Theta)$  except for the empty set and  $p_{\theta,j}$  is the belief degree.  $(\theta, p_{\theta,j})$  is a focal element of evidence  $e_j$  and represents that the evidence points to proposition  $\theta$ , to the degree of  $p_{\theta,j}$ . Both global ignorance and local ignorance are taken into account in the definition of evidence.

The ER rule enhances Dempster's rule by introducing evidence reliability and conducting reliability perturbation analysis for combining multiple pieces of evidence that are fully reliable but highly conflicting with each other.

A piece of evidence  $e_j$  is characterized by three elements including a belief distribution  $(\theta, p_{\theta,j})$ , reliability  $r_j$  and weight  $w_j$  in the framework of the ER rule. The reliability of evidence is an inherent property and measures the true rate of a proposition when the evidence points to the proposition. The weight of evidence is used to reflect its relative importance in comparison with other evidence. A weighted belief distribution with reliability (WBDR) is defined as follows, in order to combine the evidence with the above characteristics.

$$m_j = \{(\theta, \tilde{m}_{\theta,j}), \forall \theta \subseteq \Theta; (P(\Theta), \tilde{m}_{P(\Theta),j})\}$$

where  $\tilde{m}_{\theta,j}$  represents the degree of support for proposition  $\theta$  from evidence  $e_j$  while taking both the weight  $\omega_j$  and reliability  $r_j$  into consideration, defined as follows

$$\tilde{m}_{\theta,j} = \begin{cases} 0 & \theta = \phi \\ c_{rw,j} m_{\theta,j} & \theta \subseteq \Theta, \theta \neq \phi \\ c_{rw,j} (1 - r_j) & \theta = P(\Theta) \end{cases}$$

$$c_{rw,j} = 1/(1 + w_j - r_j), m_{\theta,j} = \omega_j p_{\theta,j}$$

where  $c_{rw,j}$  is a normalization factor uniquely determined to satisfy  $\sum_{\theta \subseteq \Theta} \tilde{m}_{\theta,j} + \tilde{m}_{P(\Theta),j} = 1$  given  $\sum_{\theta \subseteq \Theta} p_{\theta,j} = 1$ .  $\tilde{m}_{P(\Theta),j}$  is the degree of residual support and reflects the unreliability of the evidence  $e_j$ .

The ER rule implements the orthogonal sum operation on WBDRs and constitutes a generic conjunctive probabilistic reasoning process to combine multiple pieces of independent evidence with various weights and reliabilities.

### 3. Methodology

In this study, we developed a classifier based on the ER rule as a core component of a decision support system for handling customer complaints. The architecture of the evidential reasoning-based DSS is designed based on web technologies, as shown in Fig. 1. A three-layer intelligent DSS is developed which includes a client layer, an application server layer and a database server layer. Three components, namely a web-based user interface, an ER rule-based classifier and a database are designed in the DSS. As a core component, the ER rule-based classifier is designed to classify customer complaints before a resolution is communicated to customers. Five key steps in the classifier are feature extraction and selection, evidence acquisition for each feature, weight calculation for each piece of evidence, evidence combination and reliability training. The input of the classifier are the features representing customer complaints, and the output is the estimated belief degrees with which each customer complaint is assigned to a pre-defined class.

#### 3.1. Feature extraction and selection

The features used to characterize a customer complaint can be classified into two categories, e.g., numeric and textual features. The narratives of customer complaints usually contain information about the signal strength and call quality, which provide important signs about the likely cause of failures. Textual features are extracted firstly in this study to characterize a customer complaint. Then the most important features among them are selected together with numeric features for decision making. The numeric features are obtained from business support systems where the operation states of telecommunication networks are monitored.

A vector space model is used to represent the textual features of a customer complaint. It is a vector of features with weights [16]. Features are formed from the keywords in a complaint description and weights are represented as a term frequency-inverse document frequency (TF-IDF) which value is between zero and one. Two steps are included to obtain the vector space model of customer complaints. The first step is to extract textual features using the bag of words (BOW) method which has been widely used to produce  $n$ -gram text features [17]. Simple features are extracted to represent textual information in the forms of unigrams, bigrams and trigrams. It means that the simple features obtained from the BOW method are in the form of some keywords or phrases with one, two or three words. Then the TF-IDF method

is used for feature weighting. The term frequency (TF) of each word expresses the importance of the word in a document while the inverse document frequency (IDF) of each word represents the importance of each word in a document database. The *TF-IDF* is a statistic value that can be used to reflect how important a word is to a document in a collection. It has been successfully used for estimating the term weights to select the important individual terms [18]. Therefore, a set of textual features are extracted to represent customer complaints as the forms of unigrams, bigrams or trigrams. The feature values for each complaint are the *TF-IDF* values of these words.

Feature selection is required since the size of the initial textual features extracted from complaint descriptions is large and some features may be irrelevant to final classification results. The value of the information gain ratio (IGR) of features is used for feature selection, in order to reflect the impact of the intrinsic values of features. Suppose that a feature variable  $X$  is related to a category variable  $Y$ . The feature with high IGR values is regarded as being highly relevant to the category variable  $Y$ . The IGR is a ratio of information gain (IG) to the intrinsic value (IV). That is:

$$IGR(Y, X) = \frac{IG(Y, X)}{IV(X)} = \frac{H(Y) - H(Y|X)}{IV(X)}$$

where  $H(Y) = -\sum_{y \in Y} p(y) \log_2 p(y)$  is the entropy of variable  $Y$  and  $p(y)$  is the marginal probability density function of  $Y$ . The entropy is a measure commonly used in the information theory to characterize the uncertain of a collection of samples. The less the entropy is, the less the uncertainty is.  $H(Y|X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log_2 p(y|x)$  is the entropy of  $Y$  after observing variable  $X$  where  $p(y|x)$  is the conditional probability of  $y$  given  $x$ . The information gain  $IG(Y, X)$  is defined as a measure to reflect the decreased amount of the entropy of variable  $Y$  when the feature  $X$  is added [19]. It measures the contributions of the feature  $X$  by calculating the  $IG$  of  $Y$  after adding the feature  $X$ . The intrinsic value of the feature  $X$  is denoted as  $IV(X) = -\sum_{x \in X} p(x) \log_2 p(x)$ . In sum, the information gain ratio can reduce a bias towards the multi-valued features by taking the number and size into account when selecting a feature  $X$  to predict variable  $Y$ .

### 3.2. Evidence acquisition for each feature

The features selected above are used to characterize a customer complaint, which can be denoted as  $x_i$  ( $i = 1, \dots, M$ ) where  $M$  is the total number of selected features. Suppose that a frame of discernment  $\Theta = \{y_1, y_2, \dots, y_N\}$  is a collection of mutually exclusive and collectively exhaustive classes to which customer complaints may belong to. A belief distribution (BD) function is used to measure the extent to which the value of a feature points to each class and the subsets of the classes. Such a belief distribution is also called a piece of evidence. It can be obtained through Bayesian statistics from data on features such as past complaints and classification results [20].

#### (1) Determining the referential values of features

Suppose that a feature  $x_i$  ( $i = 1, \dots, M$ ) is a discrete variable and it has  $L$  referential values  $A_j^i$  ( $i = 1, \dots, M, j = 1, \dots, L$ ). Then the relationship between the feature  $x_i$  and the class  $y_n$  is transformed into the relationships between the referential value  $A_j^i$  ( $i = 1, \dots, M, j = 1, \dots, L$ ) of the feature  $x_i$  ( $i = 1, \dots, M$ ) and the class  $y_n$  ( $n = 1, \dots, N$ ). If the feature  $x_i$  ( $i = 1, \dots, M$ ) is a continuous variable, its referential values  $A_j^i$  can be given by experts or trained from data.

#### (2) Calculating likelihood in Bayesian paradigm

Likelihood is used to represent the probability of the referential value  $A_j^i$  ( $i = 1, \dots, M, j = 1, \dots, L$ ) for a given class  $y_n$  ( $n = 1, \dots, N$ ), as shown in Table 1. It is often derived from observed

**Table 1**  
Likelihood of feature  $x_i$ .

	$A_1^i$	...	$A_j^i$	...	$A_L^i$
$y_1$	$c_{11}$		$c_{1j}$		$c_{1L}$
...					
$y_n$	$c_{n1}$		$c_{nj}$		$c_{nL}$
...					
$y_N$	$c_{N1}$		$c_{Nj}$		$c_{NL}$

**Table 2**  
The belief degrees of the feature  $x_i$ .

	$A_1^i$	...	$A_j^i$	...	$A_L^i$
$y_1$	$p_{1,1}$		$p_{1,j}$		$p_{1,L}$
...					
$y_n$	$p_{n,1}$		$p_{n,j}$		$p_{n,L}$
...					
$y_N$	$p_{N,1}$		$p_{N,j}$		$p_{N,L}$

data in Bayesian paradigm. Let  $c_{nj}$  ( $n = 1, \dots, N, j = 1, \dots, L$ ) stand for the likelihood to which the feature  $x_i$  ( $i = 1, \dots, M$ ) is identified as the referential value  $A_j^i$  given the known class  $y_n$  ( $n = 1, \dots, N$ ). That is,

$$c_{nj} = p(A_j^i|y_n), \sum_{j=1}^L c_{nj} = 1 \quad \text{for } n = 1, \dots, N \quad (1)$$

#### (3) Transforming likelihood to evidence

A piece of evidence in the ER paradigm is used to represent the relationship between the referential value  $A_j^i$  ( $i = 1, \dots, M, j = 1, \dots, L$ ) of the feature  $x_i$  ( $i = 1, \dots, M$ ) and the class  $y_n$  ( $n = 1, \dots, N$ ). Prior research has proved that ER rule and Bayes rule become equivalent when the reliability and weight of evidence are both equal to 1 and the belief degrees are assigned to only singleton classes [20]. Let  $p_{n,j}$  ( $n = 1, \dots, N, j = 1, \dots, L$ ) stands for the belief degree that a customer complaint is belonged to the class  $y_n$ . It is equal to the normalized likelihood in Bayesian paradigm. Therefore, the evidence for each feature is acquired from sample data by normalizing likelihood.

Suppose that the likelihood above is obtained from records of the complaints made by customers independently, the relationship between the belief degree  $p_{n,j}$  in Table 2 and the likelihood  $c_{nj}$  in Table 1 is given by

$$p_{n,j} = c_{nj} / \sum_{k=1}^N c_{kj}, \sum_{n=1}^N p_{n,j} = 1, \text{ for } j = 1, \dots, L \quad (2)$$

Therefore, a piece of evidence  $e_j$  can be profiled as follows

$$e_j = \left\{ (y_n, p_{n,j}), n = 1, \dots, N, \sum_{n=1}^N p_{n,j} = 1 \right\}, j = 1, \dots, L$$

It indicates that a feature with  $L$  referential values can generate  $L$  pieces of evidence. For each referential value of a feature, there exists a piece of evidence.

### 3.3. Calculating of the weight of evidence

In the ER rule, the reliability and weight of evidence are defined separately. Shafer firstly analyzed the concept of evidence weight in an intuitive and straightforward manner [21]. It is stated that the effect of evidence is limited to providing a certain degree of support for a single non-empty subset. A discounting factor is proposed and interpreted as the reliability or weight of the evidence, depending on circumstances in which the factor is used. Smarandache et al. [22] firstly distinguished the notion between reliability and weight, and proposed a new approach for combining sources of evidences

with different importance and reliabilities in information fusion process. The weight of a source is denoted by an important factor  $\beta$  and  $0 \leq \beta \leq 1$ . In the ER approach for multicriteria decision problems, the assessment of an alternative on each criterion is regarded as a piece of evidence and the weight of evidence is equal to the weight of a criterion [23].

Here, the weight of evidence is represented by the classification capability of its corresponding features for classification outcomes. It reflects the relative importance of the evidence which provides the correct classification results according to its corresponding features. The more important the evidence is, the more contribution it makes to classification results.

We define a uniform distribution  $e_\gamma = \{(y_n, p_N), n = 1, 2, \dots, N\}$  that represents a piece of evidence obtained from its related feature values. It means that a customer complaint is believed to belong to the class  $y_n$  with the same belief degree  $p_N$  ( $p_N = 1/N$ ) when judged by a certain feature.

Let  $d_{j\gamma}$  stands for the Euclidean distance between the evidence  $e_j$  and the uniform distribution  $e_\gamma$ .

$$d_{j\gamma} = d(e_j, e_\gamma) = \sqrt{\sum_{n=1}^N (p_{n,j} - p_N)^2}, j = 1, \dots, L, d_{j\gamma} \in [0, 1] \quad (3)$$

The larger the distance is, the higher the importance of the evidence  $e_j$  is. When  $N$  tends to be infinite,  $d_{j\gamma}$  infinitely approaches 1.

### 3.4. Evidence combination using the ER rule

The value of features can be nominal or numeric. Suppose that the  $M$  features used to characterize a complaint  $x$  are nominal, each feature  $x_i$  ( $i = 1, \dots, M$ ) has  $L$  referential values  $A_j^i$  ( $j = 1, \dots, L$ ) and there are  $N$  possible classes that the complaint may belong to. There are totally  $M \cdot L$  pieces of evidence. A certain sample of customer complaint  $x$  will activate  $M$  pieces of evidence  $e_i$  ( $i = 1, \dots, M$ ).

Otherwise, for a certain sample of customer complaint  $x = (x_1, \dots, x_i, \dots, x_M)$  with numeric features, the feature  $x_i$  will activate two adjacent pieces of evidence  $e_j^i$  with the activated weight  $\alpha_{ij}$  and  $e_{j+1}^i$  with the activated weight  $\alpha_{i,j+1}$  if the feature  $x_i$  takes its value in the interval  $[A_j^i, A_{j+1}^i]$ . The following information transformation technique is used to generate the activated weights [24].

$$\alpha_{i,j} = \frac{A_{j+1}^i - x_i}{A_{j+1}^i - A_j^i} \text{ and } \alpha_{i,j+1} = 1 - \alpha_{i,j},$$

$$\text{if } A_j^i \leq x_i \leq A_{j+1}^i, i = 1, \dots, M, j = 1, \dots, L$$

$$\alpha_{i,j'} = 0, \text{ for } j' = 1, \dots, L, j' \neq j, j + 1$$

Therefore, the evidence  $e_i$  of the feature  $x_i$  with the belief degree  $p_{ni}$  can be calculated as the weighted sum of evidence  $e_j^i$  and  $e_{j+1}^i$  as follows:

$$e_i = \{(y_n, p_{n,i}), n = 1, \dots, N, i = 1, \dots, M\}$$

$$\text{where } p_{n,i} = \alpha_{i,j} p_{n,j} + \alpha_{i,j+1} p_{n,j+1}, j = 1, \dots, L$$

Here,  $p_{ni}$  is the belief degree to which the class is believed to be  $y_n$  given that the feature  $x_i$  activates  $e_j^i$  and  $e_{j+1}^i$  simultaneously. Meanwhile, the weight  $w_i$  and reliability  $r_i$  of the evidence  $e_i$  can be calculated as the weighted sum of those of the two adjacent pieces of evidence. That is:

$$w_i = \alpha_{i,j} w_j^i + \alpha_{i,j+1} w_{j+1}^i \text{ and } r_i = \alpha_{i,j} r_j^i + \alpha_{i,j+1} r_{j+1}^i,$$

$$i = 1, \dots, M, j = 1, \dots, L$$

Then the ER rule is used to combine the activated evidence with their corresponding weights and reliabilities. The fused result  $O(x)$  is as follows,

$$O(x) = \{(y_n, p_{n,e(M)}), n = 1, \dots, N\}$$

where the certain sample  $x$  belongs to a specific class  $y_n$  that has the maximum belief degree  $p_{n,e(M)}$ . That means if  $\max(p_{1,e(M)}, p_{2,e(M)}, \dots, p_{N,e(M)})$  equals to  $p_{i,e(M)}$ , the fusion result of the  $M$  pieces of evidence point to the class  $y_i$ . The combined belief degree,  $p_{n,e(M)}$ , is denoted as follows,

$$p_{y,e(M)} = \begin{cases} 0 & n = \emptyset \\ \frac{\hat{m}_{y,e(M)}}{\sum_{B \subseteq \Theta} \hat{m}_{B,e(M)}} & n \neq \emptyset \end{cases} \quad (4)$$

where  $\hat{m}_{y,e(i)} = [(1 - r_i) m_{y,e(i-1)} + m_{p(\Theta),e(i-1)} m_{y,i}] + \sum_{B \cap C = y} m_{B,e(i-1)} m_{C,i}, \forall y \subseteq \Theta$  and  $\hat{m}_{p(\Theta),e(i)} = (1 - r_i) m_{p(\Theta),e(i-1)}, i = 1, \dots, M$ . It represents the degree to which  $M$  pieces of independent evidence  $e_i$  ( $i = 1, \dots, M$ ) with weights  $w_i$  ( $i = 1, \dots, M$ ) and reliabilities  $r_i$  ( $i = 1, \dots, M$ ) jointly support proposition  $y_n$ .

### 3.5. Training the reliability of evidence

It has been discussed above that there is a distinction between the weight and the reliability of evidence in the ER rule paradigm. The former represents the relative importance of evidence compared with other evidence while the latter measures the ability of evidence to provide correct assessment or solution for a given problem. Generally speaking, the weight of evidence shares the same definition as that of its reliability if all pieces of evidence are obtained from the same set of data [14]. In many real-life information infusion problems, however, different sources of information arise from human experts, systems or physical sensors. The sources may not have the same reliability, neither the same importance [22].

In the proposed classification model, each piece of evidence is obtained from likelihood that generated from data and its related evidential elements are associated with a subset of categories. Evidence reliability is used to express the limited power of these evidential elements to assert the subset. It can be equal to evidence weight initially and then be acquired by optimal learning from data while minimizing the difference between the real output and the fused outcome through the ER rule-based classifier.

Suppose that a dataset  $S$  including  $K_s$  customer complaints is used to train the reliability  $r_i$  ( $i = 1, \dots, M$ ) of evidence  $e_i$  ( $i = 1, \dots, M$ ). For a certain complaint  $x^k$  in the dataset  $S$ , there is a corresponding class  $y_k$  that  $x^k$  actually belongs to. Suppose that the fused result of the classifier for input  $x^k$  is  $O(x^k) = \{(y_n, p_{n,e(M)}), n = 1, \dots, N\}$ . An optimization model based on the Euclidean distance is developed as follows.

$$\min \xi(p) = \frac{1}{K_s} \sum_{k=1}^{K_s} d_E(u^k, v^k) \quad (5)$$

$$\text{s.t. } 0 \leq r_i \leq 1, i = 1, \dots, M$$

where  $p = \{r_i | i = 1, \dots, M\}$  stands for the parameters to be optimized.  $u_k$  is the belief degree vector form of  $O(x^k)$  and  $v_k$  is that of the class  $y_k$ .

Take a two-class problem for example. If the complaint  $x^k$  belongs to the class  $y_1$  and the fused outcome through the ER rule-based classifier is  $O(x^k) = \{(y_1, p_{1,e(M)}), (y_2, p_{2,e(M)})\}$ , then  $v_k = (1, 0)$  and  $u_k = (p_{1,e(M)}, p_{2,e(M)})$ . A non-linear programming method will be introduced to solve this optimization problem. When the objective function  $\xi(p)$  is minimized, all the reliabilities of evidence reaches to an optimal value.

## 4. An empirical study

In this section, an empirical study is conducted with a telecommunication network operator where customer complaints related to network quality are processed by experienced technicians. Owing to the complexity of telecommunication technology, there is

**Table 3**

A sample of a customer complaint narrative.

Description
Time of failure: 2015-08-31 22:36
Location: Longfeng community, Qianshan Bei Road, Shushan District
Customer's description: It is particularly slow to access the Internet by the mobile phone. The signal is bad and the number of signal strength dots is two.
On-off testing: No
Signal of surrounding users: not clear

much uncertain or conflicting information when customers complain about service or product failure. The inconsistent information in customer complaint narratives greatly influences the technicians when they make a judgment on complaint causes. Therefore, we developed an ER-based decision support system to improve the efficiency of handling complaints.

4.1. Problem description

There are many complaints about the telecommunication services from the subscribers of the surveyed mobile network operator. Different complaints on specific services will be classified and assigned to a technical support department for resolution. A typical complaint handling process in mobile telecommunications is described in Fig. 2. When the receptionists in customer service center receive a customer complaint, they record it and make a judgment whether it is a valid complaint by their experience or using CRM systems. If they can handle the complaint and resolve it on their own independently, they will send feedback directly to the customer. Otherwise, they will assign the complaint to appropriate service support departments. Technicians in the departments will analyze the cause of complaints, provide solutions and send feedback to customers. When making judgment about the cause of complaints, technicians need to make good use of both the information from business support systems and the narratives of customer complaints. Then the complaints will be classified into two categories including network quality class (NQC) and customer terminal class (CTC) according to the cause of complaints. Owing to the obtained information is often inconsistent, it is difficult for technicians to make accurate judgments, thus resulting in classification imprecision.

4.2. Data collection

We collected the customer complaint narratives recorded during the period from August to October in 2015 from a technical support department in the surveyed telecommunication company. In total, 1433 records are obtained as sample data after eliminating some invalid complaints. The sample data include details about customer complaints, handling results, and corresponding feedback to customers. A typical complaint record is listed in Table 3, which includes the time and location of service failure, the customer's description about the perception of network services, such as slow internet access and poor call quality.

Meanwhile, we have interviewed a group of professional technicians who specialize in handling complaints. Six numeric features are also selected from business support systems in the surveyed company, which are as follows: (1) *D*—whether the complaint is from an area where signal interference is in force; (2) *F*—whether the complaint is from a crowd-gathering area; (3) *CT*—whether the complaint happens in the period of crowd gathering; (4) *BS*—whether the operational state of the current base station is healthy; (5) *B*—whether the complaint is from an area with weak signal strength; (6) *MP*—whether there is a compatibility issue between the mobile telecommunication network and the terminal used by the complaining customer. All the features are represented

**Table 4**

Top ten textual features selected from complaint descriptions.

Features ID	Chinese	English	Information gain ratio
F1	晚上	Evening	0.0108
F2	网页	Webpages	0.0104
F3	功能	Function	0.0103
F4	打开	Open	0.0102
F5	只能	Only	0.0101
F6	开通	Start	0.0100
F7	打不开	Failure	0.0098
F8	速度	Speed	0.0098
F9	数据	Data	0.0097
F10	主叫	Calling	0.0010

**Table 5**

Likelihood and belief degrees of the feature *D*.

		Referential values of <i>D</i>		
		1( <i>e</i> <sub>1</sub> )	0( <i>e</i> <sub>2</sub> )	Unknown( <i>e</i> <sub>3</sub> )
Likelihood	NQC	0.033	0.967	0
	CTC	0	1	0
Belief degree	MTNQC	1	0.4916	0
	CTC	0	0.5084	0

by three referential values {zero, one, unknown} where *zero* means *NO*, *one* is *YES*, and *unknown* represents the missing or unknown information.

4.3. Extraction and selection of textual features

Since the values of the six key features may not be enough for technicians to make an accurate decision, the textual features are extracted from complaint descriptions to characterize a complaint and then selected for classification together with the six key features above. The TF-IDF method is used to extract the textual features. According to the information gain ratio values of the features, top ten textual features are selected for classification, as shown in Table 4.

Since the TF-IDF values of these ten textual features for each complaint narrative are continuous, they will be discretized before evidence acquisition. Their referential values are set to be {0, 0.33, 0.66} according to the data interval.

4.4. Evidence acquisition

Evidence is acquired for each feature by Bayesian statistics. Take the feature *D* as an example. *D* represents whether the complaint is from an area where signal interference is in force. Since the value of the feature *D* is discrete, its referential value  $A_j^i = \{1, 0, \text{unknown}\}$ . The likelihood  $c_{nj}$  and belief degree  $p_{nj}$  of the feature *D* can be obtained using Eqs. (1) and (2), as shown in Table 5. Three pieces of evidence related to the feature *D* can then be profiled as follows:

$$e^D_1 = \{(NQC, 1), (CTC, 0)\}, e^D_2 = \{(NQC, 0.4916), (CTC, 0.5084)\}, e^D_3 = \{(NQC, 0), (CTC, 0)\}.$$

It means that if the value of the feature *D* of a complaint equals to 1, the probability that the complaint belongs to NQC is 100%. Otherwise, the complaint belongs to NQC with the probability of 49.16% and to CTC with that of 50.84%.

If the information of features is incomplete due to missing data in records, there will be ignorance in calculating likelihood. In the scheme of the ER rule, the ignorance is represented by a belief degree assigned to the attribute with the value “unknown”. Take the feature *MP* for example. Its referential value  $A_j^i$  is one of the

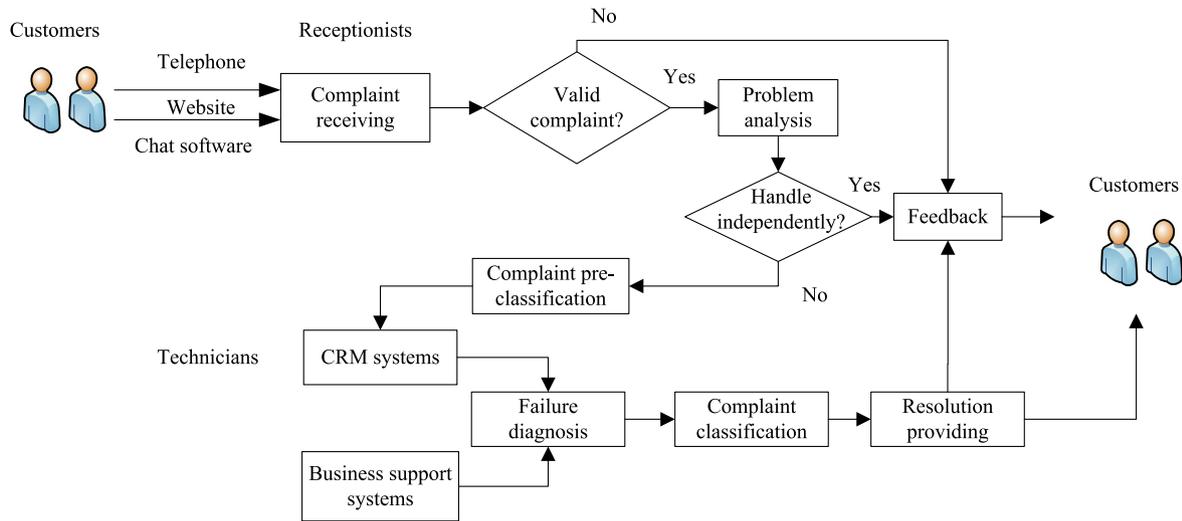


Fig. 2. A typical complaint handling process in mobile telecommunications.

Table 6  
The evidence and its weights.

Features	Referential values	Evidence	NQC	CTC	Weights
D	1	$e^{D_1}$	1.0000	0.0000	0.8498
	0	$e^{D_2}$	0.4900	0.5100	0.4716
	Unknown	$e^{D_3}$	0.0000	0.0000	0.2357
F	1	$e^{F_1}$	0.2360	0.7640	0.6014
	0	$e^{F_2}$	0.6309	0.3691	0.5065
	Unknown	$e^{F_3}$	0.0000	0.0000	0.2357
CT	1	$e^{CT_1}$	0.0000	1.0000	0.8498
	0	$e^{CT_2}$	0.6794	0.3206	0.5353
	Unknown	$e^{CT_3}$	0.0000	0.0000	0.2357
BS	1	$e^{BS_1}$	0.4911	0.5089	0.4716
	0	$e^{BS_2}$	0.9003	0.0997	0.7367
	Unknown	$e^{BS_3}$	0.0000	0.0000	0.2357
B	1	$e^B_1$	0.9600	0.0400	0.8034
	0	$e^B_2$	0.3293	0.6707	0.5296
	Unknown	$e^B_3$	0.0000	0.0000	0.2357
MP	1	$e^{MP_1}$	0.6826	0.3174	0.5375
	0	$e^{MP_2}$	0.2439	0.7561	0.5944
	Unknown	$e^{MP_3}$	0.0971	0.9029	0.7395

value in the set {1,0, unknown}, and the evidence related to each of the referential values of the feature is as follows.

$$e^{MP_1} = \{(NQC, 0.6826), (CTC, 0.3174)\},$$

$$e^{MP_2} = \{(NQC, 0.2439), (CTC, 0.7561)\},$$

$$e^{MP_3} = \{(NQC, 0.0971), (CTC, 0.9029)\}$$

#### 4.5. Elicitation of weight and reliability

We define a distribution  $e_\gamma = \{(y_n, p_n)\}, n = 1, 2, \dots, 16\}$  that represents a piece of evidence without any classification capability. Since 10 textual and 6 numeric features in total are used for classification,  $p_n$  is a constant with the value of 1/16. The weight of each piece of evidence is obtained by Eq. (3). The evidence related to six numeric features and its weights are listed in Table 6.

The reliability of evidence is set initially as the same as its weight in the classification context and then is trained through learning from sample data. The *Fmincon* function in Matlab's optimization toolbox is used to seek the optimal reliability of each

Table 7  
A confusion matrix.

	The predicted class	
	NQC	CTC
Actual class of complaints	NQC	True positive (TP)
	CTC	False positive (FP)
		False negative (FN)
		True negative (TN)

piece of evidence while minimizing the Euclidean distance between the real output and the fused output of the ER rule-based classifier.

#### 4.6. Performance measures

Once a customer complaint is fed into the evidential reasoning-based decision support system, each piece of evidence related to features is activated and fused by the recursive analytical model of the ER rule. We use a confusion matrix to measure the performance of the ER rule-based classifier. The confusion matrix is a table with two rows and two columns that represents the number of *true positive* (TP), *false positive* (FP), *false negative* (FN) and *true negative* (TN), as shown in Table 7.

If a complaint which actually belongs to NQC is predicted as NQC, it is counted as a *true positive*; if it is classified as CTC, it is counted as *false negative*. If the complaint which actually belongs to CTC is classified as NQC, it is counted as *false positive*; if it is classified as CTC, it is counted as a *true negative* [25]. Four performance indicators including accuracy, recall (or sensitivity), precision and F-measure are calculated in the experiment through ten-fold cross validation.

$$Accuracy = \frac{TP + TN}{Total}, recall = \frac{TP}{TP + FN}, precision = \frac{TP}{TP + FP},$$

$$F\text{-measure} = \frac{2}{\frac{1}{precision} + \frac{1}{recall}}$$

Accuracy is to measure how often the classifier is correct. Recall, known as sensitivity or true positive rate, is to measure how often the classifier predicted NQC when the customer complaint actually belongs to NQC. Precision, also called positive predictive value, is the number of NQC divided by the total number of elements labeled as belong to the NQC. F-measure (or F1 score) is a measure of the accuracy of the study with both the precision and the recall taken into consideration.

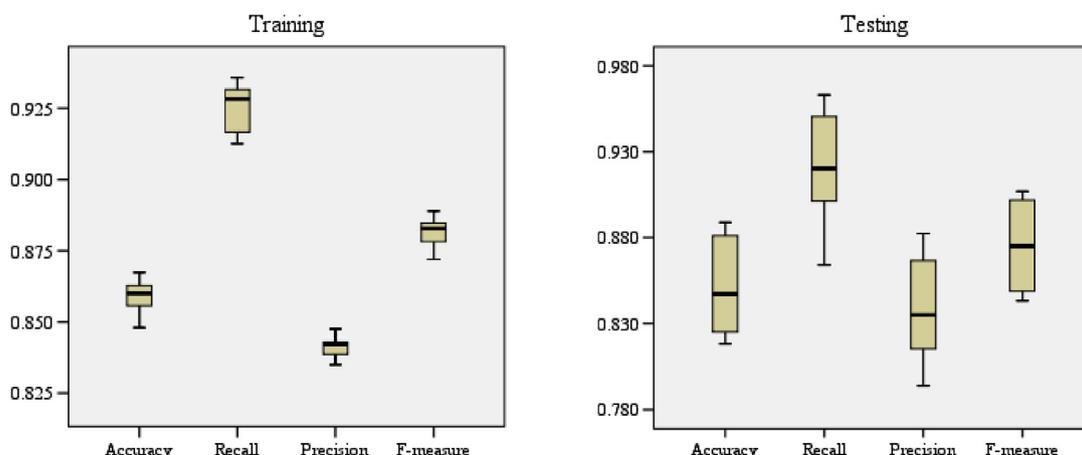


Fig. 3. The box plot of performance measures in ten-fold cross-validation.

Table 8

The average values of ten-fold cross-validation results.

	Accuracy	Precision	Recall	F-Measure
Training	0.8591	0.9259	0.8416	0.8817
Testing	0.8527	0.9188	0.8383	0.8762

Ten textual and six numeric features are selected for the experiment. Ten-fold cross-validation is conducted to evaluate the reliability of the proposed classification model. The model is trained and tested ten times, respectively. The averages of the ten-fold cross-validation results on the four performance measures described above are shown in Table 8. Meanwhile, the ten-fold cross-validation results of both training and testing are graphically depicted as a box plot in Fig. 3, where five values including minimum, first quartile, median, third quartile and maximum are used to measure the distribution of ten results. Undoubtedly, the full range of variation of the testing results is greater than that of the training results.

### 5. Discussions

The classification performance of the proposed ER rule-based classifier may vary with the number of features. To enhance the predictive ability of the classification model, the performance of the model with different textual features and numeric features are analyzed below. The comparisons of performance measures are shown in Fig. 4. The number of textual features is changed from ten to fifty for comparison. Experiments show that the classification performance of the sixteen features achieves an overall high performance. It also indicates that it is necessary to extract textual features from customer descriptions for decision-makers although there is uncertain and conflicting information in the complaint narratives.

Classical probabilistic modeling tools including Logistic regression (LR), Bayesian networks (BNs) and Naïve Bayes (NBs), and non-probabilistic methods including support vector machine (SVM) and J48 decision tree are often deployed to construct classifiers. Also, the typical ensemble classifiers including random forest, bagging and adaboosting are adopted for calculation. To compare their classification performance with that of the proposed ER rule-based classifier, all the nine classifiers are trained using the same dataset. Except for the ER rule-based classifier implemented in Matlab software, the other eight classifiers are developed in WEKA software, an open-source data mining toolkit. The parameters of the learning

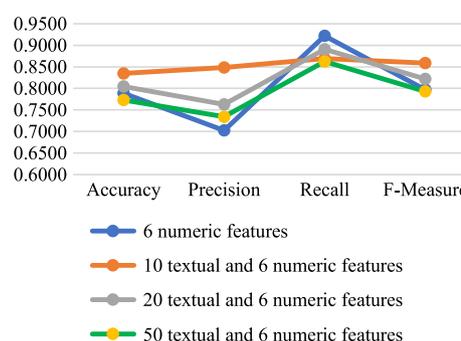


Fig. 4. The classification performance with different numbers of features.

Table 9

The performance of nine classifiers with 16 features.

	Accuracy	Precision	Recall	F-measure
The ER rule-based classifier	0.8559	<b>0.9223</b>	0.8400	<b>0.8790</b>
Bayes Net	0.8570	0.8570	0.8600	0.8550
Naïve Bayes	0.8530	0.8530	0.8570	0.8520
Logistic regression	0.8420	0.8420	0.8420	0.8420
SVM	0.8000	0.8000	0.8440	0.7870
J48 decision tree	<b>0.8590</b>	0.8590	<b>0.8750</b>	0.8550
Bagging	0.8460	0.8460	0.8570	0.8420
Adaboost	0.8300	0.8300	0.8310	0.8280
Random forest	0.8540	0.8540	<b>0.8750</b>	0.8490

algorithms of the eight classifiers all take default values in the WEKA software.

The performance of the nine classifiers with 16 features are as shown in Table 9. It indicates that the ER rule-based classifier has competitive performance against other classifiers when both textual and numeric features are used for classification. There is a little enhancement in the precision values of the ER rule-based classifier in comparison of both the non-probabilistic and ensemble classifiers.

It can be seen from the comparative analysis that the proposed ER-based DSS has advantages below. Firstly, it provides a data-driven and probabilistic modeling tool when the evidence and its weight and reliability are all formulated from real data. Secondly, it can aggregate inconsistent evidence under uncertainty for classification. A piece of evidence is profiled by a belief distribution on the power set of the frame of discernment. The natural and flexible form of probability distribution allows inexact reasoning at whatever level of abstraction. Lastly, it can be used to combine

multi-source heterogeneous data for decision-making. One of the possible limitations is that both structured data from business support systems and unstructured data from complaint narrations have to be transformed to evidence for decision-making.

## 6. Conclusions

In this paper, an evidential reasoning-based decision support system is developed for handling customer complaints in mobile telecommunications. Uncertain information is formulated as a belief distribution and the roles of different features are distinguished through their corresponding evidence with different reliabilities and weights. The proposed evidential reasoning rule-based classifier is equipped with the inherent features for handling missing data without deletion or imputation.

An empirical study is conducted in a leading telecommunication operator, where six numeric and ten textual features related to telecommunication quality are selected to characterize a customer complaint in mobile telecommunications. A systematic process of handling complaints is described and the input–output of the proposed classifier is distinguished. The experimental results show its high classification accuracy that is competitive with classical probabilistic models including Bayes net and Naïve Bayes, and some non-probabilistic models, such as SVM and J48 decision tree. Moreover, the performance of the proposed ER rule-based classifier is superior to some ensemble classifiers such as bagging and adaboosting classifiers. The proposed web-based intelligent DSS provides telecommunication technicians with an informative and knowledge-based methodology for handling customer complaints systematically and automatically. It is helpful to facilitate handling customer complaints efficiently and effectively so as to further improve customer satisfaction.

However, binary classification is only taken into consideration in the empirical study. Multiple-class classification problems should be further investigated in the future. Moreover, we only used the conventional optimization toolbox in MATLAB for solving the optimization model of learning the reliability of evidence, and state-of-the-art metaheuristic algorithms can be further explored to improve the computation efficiency.

## Acknowledgments

This research was supported by the National Nature Science Foundation of China (Nos. 71573071, 71571060, 71671057, 71771077).

## References

- [1] T. Garín-Muñoz, T. Pérez-Amaral, C. Gijón, et al., Consumer complaint behavior in telecommunications: The case of mobile phone users in Spain, *Telecommun. Policy* 40 (8) (2016) 804–820.
- [2] B.A. Galitsky, M.P. Gonzalez, C.I. Chesnevar, et al., A novel approach for classifying customer complaints through graphs similarities in argumentative dialogues, *Decis. Support Syst.* 46 (3) (2009) 717–729.
- [3] K. Coussement, D.V.D. Poel, Improving customer complaint management by automatic email classification using linguistic style features as predictors, *Decis. Support Syst.* 44 (4) (2008) 870–882.
- [4] A. Faed, O.K. Hussain, E. Chang, A methodology to map customer complaints and measure customer satisfaction and loyalty, *Serv. Oriented Comput. Appl.* 8 (1) (2014) 33–53.
- [5] C. Lee, Y. Wang, A.J. Trappey, et al., Ontology-based reasoning for the intelligent handling of customer complaints, *Comput. Ind. Eng.* (2015) 144–155.
- [6] A. Faed, E. Chang, M. Saberi, et al., Intelligent customer complaint handling utilising principal component and data envelopment analysis, *Appl. Soft Comput.* 47 (C) (2016) 614–630.
- [7] F. Browne, N. Rooney, W. Liu, et al., Integrating textual analysis and evidential reasoning for decision making in engineering design, *Knowl.-Based Syst.* 52 (6) (2013) 165–175.
- [8] Y. Ko, H. Fujita, T. Li, et al., An evidential analysis of altman z-score for financial predictions, *Appl. Soft Comput.* (2017) 748–759.
- [9] Y. Ko, H. Fujita, Evidential probability of signals on a price herd prediction: Case study on solar energy companies, *Int. J. Approx. Reason.* (2018) 235–269.
- [10] X. Xu, J. Zheng, J.B. Yang, et al., Data classification using evidence reasoning rule, *Knowl.-Based Syst.* 116 (2017) 144–151.
- [11] B. Jugwanth, D. Vigar-Ellis, Customer complaint behaviour and companies' recovery initiatives: the case of the hello peter website, *Ellis* 20 (2013) 143–165.
- [12] R. Johnston, Linking complaint management to profit, *Int. J. Serv. Ind. Manag.* 12 (1) (2001) 60–69.
- [13] K. Jeschke, H.S. Schulze, J. Bauersachs, Internal marketing and its consequences for complaint handling effectiveness, in: *Relationship Marketing*, Springer Berlin Heidelberg, 2000, pp. 193–216.
- [14] J.B. Yang, D.L. Xu, Evidential reasoning rule for evidence combination, *Artificial Intelligence* 205 (205) (2013) 1–29.
- [15] J. Gordon, E.H. Shortliffe, A method for managing evidential reasoning in a hierarchical hypothesis space, *Artificial Intelligence* 59 (3) (1985) 323–357.
- [16] C.D. Manning, P. Raghavan, H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, New York, NY, USA, 2008.
- [17] Y. Zhang, Y. Dang, H. Chen, et al., Automatic online news monitoring and classification for syndromic surveillance, *Decis. Support Syst.* 47 (4) (2009) 508–517.
- [18] T. Mori, A term weighting method based on information gain ratio for summarizing documents retrieved by IR systems, *J. Nat. Lang. Process.* (2002) 9.
- [19] G. Wang, J. Ma, S. Yang, An improved boosting based on feature selection for corporate bankruptcy prediction, *Expert Syst. Appl.* 41 (5) (2014) 2353–2361.
- [20] J.B. Yang, D.L. Xu, A study on generalizing bayesian inference to evidential reasoning, in: *International Conference on Belief Functions: Theory and Applications*, Springer-Verlag New York, Inc, 2014, pp. 180–189.
- [21] G. Shafer, A mathematical theory of evidence, *Technometrics* 20 (1) (1978) 106–106.
- [22] F. Smarandache, J. Dezert, J.M. Tacnet, Fusion of sources of evidence with different importance and reliabilities, in: *Information Fusion*, IEEE, 2011, pp. 1–8.
- [23] J.B. Yang, M.G. Singh, An evidential reasoning approach for multiple-attribute decision making with uncertainty, *IEEE Trans. Syst. Man Cybern.* 24 (1) (1994) 1–18.
- [24] Y. Chen, Y.W. Chen, Xu X. B., et al., A data-driven approximate causal inference model using the evidential reasoning rule, *Knowledge-Based Systems* 88 (C) (2015) 264–272.
- [25] T. Fawcett, An introduction to ROC analysis, *Pattern Recognit. Lett.* 27 (8) (2006) 861–874.