ADPDF: A Hybrid Attribute Discrimination Method for Psychometric Data With Fuzziness

Xi Xiong, Shaojie Qiao, Yuanyuan Li, Haiqing Zhang, Ping Huang, Nan Han, and Rong-Hua Li

***Abstract*—The existing approaches for attribute discrimination are applied to clinical data with unambiguous boundaries, and rarely take into careful consideration on how to utilize psychometric data with fuzziness. In addition, it is difficult for conventional attribute reduction methods to reduce attributes of psychometric data which are composed of a lot of attributes and contain a relatively small-scale samples. Importantly, these methods cannot be used to reduce options which are relevant to each other. In this paper, we first introduce new concepts, that is, option entropy and option influence degree, which are employed to describe the relation and distribution of options. Then, we propose a hybrid attribute discrimination method for psychometric data with fuzziness, called a hybrid attribute discrimination for psychometric data with fuzziness (ADPDF). ADPDF contains three essential techniques: 1) a fuzzy option reduction method, which aims to combine a fuzzy option to adjacent options, and is used to reduce the fuzziness of options in a psychometry and 2) *k*-fold attribute reduction method, which partitions all samples into several subsets and negotiates the reduction results of different subsets, and reduces the noise for the purpose of accurately discovering key attributes. In order to show the advantages of the proposed approach, we conducted experiments on two real datasets collected from clinical diagnoses. The experimental results show that the proposed method can decrease the correlation between options effectively. Interestingly, we find three reserved options and one hundred samples in each subset show the best classification performance. Finally, we compare the proposed method with typical attribute discrimination algorithms. The results reveal that our method can improve the classification accuracy with the guarantee of time performance.**

***Index Terms*—Attribute discrimination, fuzzy sets, medical data mining, option reduction.**

Manuscript received May 29, 2017; revised October 25, 2017 and January 21, 2018; accepted June 8, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61772091 and Grant 61602064, in part by the Planning Foundation for Humanities and Social Sciences of Ministry of Education of China under Grant 15YJAZH058, in part by the Youth Foundation for Humanities and Social Sciences of Ministry of Education of China under Grant 17YJCZH202, in part by the Sichuan Science and Technology Program under Grant 2018GZ0253, in part by the Innovative Research Team Construction Plan in Universities of Sichuan Province under Grant 18TD0027, in part by the Scientific Research Foundation for Advanced Talents of Chengdu University of Information Technology under Grant KYTZ201637, Grant KYTZ201715, and Grant KYTZ201750, and in part by the Scientific Research Foundation for Young Academic Leaders of Chengdu University of Information Technology under Grant J201701. This paper was recommended by Associate Editor A. Hussain. *(Corresponding author: Shaojie Qiao.)*

X. Xiong and S. Qiao are with the School of Cybersecurity, Chengdu University of Information Technology, Chengdu 610225, China (e-mail: sjqiao@cuit.edu.cn).

Y. Li is with the Mental Health Center, West China Hospital, Sichuan University, Chengdu 610041, China.

H. Zhang is with the School of Software Engineering, Chengdu University of Information Technology, Chengdu 610225, China.

P. Huang and N. Han are with the School of Management, Chengdu University of Information Technology, Chengdu 610103, China.

R.-H. Li is with the School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMC.2018.2847029

I. Introduction

IT IS widely accepted that computational intelligence has been used to in many complicated domains [1], [2]. Fuzzy analytical methods, e.g., fuzzy set theory and rough set theory, work effectively and efficiently for decision-making, and have already been applied to medical practices including the diagnoses of mental diseases [3]. The existing methods based on the fundamental theories, such as attribute reduction and fuzzy decision-making play a significant role in removing redundant information. From a medical point of view, these approaches aim to extract the most valuable information that could assist in the treatment of diseases, and may potentially reveal profound medical knowledge and provide new medical insight [4].

Mental diseases are becoming widespread around the world [5] and World Health Organization predicted in 2009 [6] that a quarter of people in the world would be affected by mental and neurological disorders in their daily lives. Psychometries [7], [8], as the primary assessment strategy for mental diseases, are used to find the probable reasons for the symptoms. A typical psychometry is a questionnaire containing many questions relevant to mental diseases. Participants need to answer these questions based on their mental health conditions. Actually, each question can be viewed as an attribute with several options to show different degrees of diseases, e.g., serious, normal, or not at all. Two illustrative examples are given as follows.

1) The revised patient health questionnaire (PHQ-15) [9] consists of 15 questions and each of them has 5 options (Table I).
2) The brief psychiatric rating scale (BPRS) [10] consists of 24 questions and each of them has 7 options (Table II).

We can see from the above examples that psychometries contain some significant information that can be used for diagnosing mental diseases.

The motivations of this paper are given as follows.

1) It is a challengeable task to distinguish unclear boundaries and fuzzy differences between options. Meanwhile,

TABLE I
PHQ-15 QUESTIONNAIRE [9] WITH 15 QUESTIONS AND 5 OPTIONS

No	Attribute	not at all	a little	moderately	a lot	severely
1	Stomach pain		✓			
2	Back pain	✓				
	...					
15	Trouble sleeping	✓				

untrained participants can hardly understand the exact meaning of some questions in the questionnaire and might choose approximate options. Thus, some attributes may be abundant and could be removed. The characteristics of fuzziness in psychiatry and psychometric data will produce noise. It is essential to reduce fuzziness before mining valuable information and extracting key features from psychometric data with fuzziness.

2) The diagnoses of mental diseases have become a complicated task and sometimes patients may be misdiagnosed by subjective psychiatrists. An expert system beyond a huge volume of psychometric data is helpful to psychiatrists or other experts to make relatively accurate decisions in diagnoses.

By carefully analyzing the psychometric data with fuzziness, we propose a hybrid attribute discrimination method to help psychiatrists or mental disease experts to diagnose more accurately and effectively. In this paper, the original contributions are given as follows.

1) We proposed two concepts namely *option entropy* and *influence degree*. The option entropy is used to describe the decisive information contained in each option and the influence degree is used to depict the relation between any two options.
2) A hybrid attribute discrimination for psychometric data with fuzziness (ADPDF), is proposed to extract key attributes from psychometry data with fuzziness. The method includes three essential steps: a) extracting data from a psychometric database; b) option reduction and overlapped attribute reduction; and c) sorting attributes and extracting key attributes.
3) The experiments are conducted on two psychometric datasets and the results show that most of the participants are apt to select parts of options instead of choosing all the options on average. Reducing and merging option can decrease the correlation of options. The operations of option reduction as well as attribute reduction can improve the classification accuracy of attribute discrimination algorithms by 3.5%–15.5% without increasing the computational complexity.

The remainder of this paper is organized as follows. Section II surveys related works in attribute reduction and machine learning approaches for diagnoses and then provides the problem statement. Section III introduces the preliminaries and important definitions. We introduce a hybrid method for key attribute discrimination in Section IV. The experimental results including the comparison with other algorithms are presented in Section V. Lastly, we conclude this paper in Section VI.

II. RELATED WORKS AND PROBLEM STATEMENT

Rough set theory and fuzzy set theory are two significant tools which have been widely applied in several research areas including pattern recognition and data mining. The main idea of rough set theory is to reduce the redundancy of data through attribute reduction [11], while preserving the ability of classification. Recently, researchers have proposed several reduction algorithms. Du and Hu [12] investigated the problem of attribute reduction for ordered decision tables based on evidence theory. Belief and plausibility functions were proposed to define relative belief and plausibility reducts of ordered decision tables. Hu *et al.* [13] proposed a theoretic framework of fuzzy-rough model based on fuzzy relations to construct a forward greedy algorithm for hybrid attribute reduction. However, the existing algorithms can only be applied to the data composed of a small number of attributes.

Another tool is fuzzy sets, in which the membership degrees are normalized between 0 and 1. Fuzzy set theory has been applied in various areas, e.g., modeling diagnostic process. Physicians' expertise was expressed by fuzzy relation of diseases and symptoms [14], and this approach has been widely used for medical diagnosis. The max–min composition method [15] and distance-based method [16] are two popular methods for medical diagnosis based on fuzzy relations [17]. The max–min composition method is intuitive. The distance-based method can decrease the loss of data information and assign weights to patients symptoms. Both methods only handle the distribution of attributes. However, the distributions of other dimensions, such as options, are totally different from attribute distribution and are not been considered. The conventional methods based on fuzzy set theory may lose some significant information for the complexity of medical data.

Researchers have proposed many machine learning methods that can imitate the human reasoning to solve problems or make decisions for diagnoses. Similar methods have also been found to deal with uncertain or incomplete information. A mental health diagnostic expert system was proposed to assist psychologists in diagnosing and treating their mental patients [18]. Various classification methods, i.e., Bayesian networks, multilayer perceptron, decision trees, single conjunctive rule learning, and fuzzy inference systems, have been applied to diagnoses of diabetes [19]. It can be observed from empirical studies that different methods yielded different accuracy levels from different accuracy measurements, e.g., Kappa statistic and error rates. A Bayesian network decision model was proposed [20] for the diagnosis of dementia and mild cognitive impairment. This model was considered to be appropriate for representing uncertainty and causality and showed better performances when compared to most of the other well-known classifiers. Multilayer perceptron with back propagation learning can diagnose Parkinson's disease effectively by a reduced number of attributes [21]. Dabek and Caban [22] proposed a neural network model with an accuracy of 82.35% for predicting the likelihood of

TABLE II
BPRS QUESTIONNAIRE [10] WITH 24 QUESTIONS AND 7 OPTIONS

No	Attribute	not present	very mild	mild	moderate	moderately severe	severe	extremely severe
1	Somatic concern						✓	
2	Anxiety				✓			
...	...							
24	Mannerisms and posturing		✓					

developing psychological conditions, such as anxiety, behavioral disorders, depression, and post-traumatic stress disorders (PTSD). Prasad *et al.* [23] proposed a hybrid architecture based on rough set theory and machine learning algorithms, which is used to predict the growth of thyroid gland diseases. Saha *et al.* [24] presented a joint modeling framework in order to classify mental health-related co-occurring online communities based on topics and psycholinguistic features expressed in the posts. Although these methods utilize typical classification algorithms for diagnose various categories of diseases including mental diseases. They might overlook some significant characteristics of psychometric data with fuzziness.

By analyzing the aforementioned attribute discrimination approaches, we can find that these algorithms focus on the clinical data with clear feature boundaries, and rarely consider how to utilize psychometric data with fuzziness. Particularly, the conventional attribute reduction methods can hardly be used to reduce attributes of a psychometric data that consist of many attributes and contain a relatively small number of samples. Moreover, these methods cannot be used to reduce options which are relevant to each other.

In order to solve these problems, we aim to achieve the following goals in this paper.

1) Due to the fuzzy characteristics of psychometries, it is significant to establish a generic framework to process the psychometric data. It can discriminate key attributes and help psychiatrists make a relatively accurate clinical diagnosis.
2) Some options seem ambiguous and have unclear boundaries. Participants can hardly distinguish them. A choice might depend on various factors, i.e., individuals' habits and moods. Many options for a question are not helpful to participants, and actually two or three options are enough for making decisions in practice. In this paper, we aim to reduce abundant options and combine them with essential options automatically.
3) We need to propose a new attribute reduction method to reduce attributes in a psychometry that contains many attributes but a relatively small number of samples. The conventional reduction methods have been proved unqualified to this task.

III. PRELIMINARIES

A. Fuzzy Option Set

Unlike the conventional set, a fuzzy set expresses the degree to which an element belongs to a set. The characteristic function of a fuzzy set is allowed to have values between 0 and 1, which denotes the degree of membership with respect to an element in a given set.

The options of one attribute shows the different degrees of symptoms in a psychometry and we define the concept of a fuzzy option set as follows.

Definition 1 (Fuzzy Option Set): Given O is a collection of options denoted generically by x, a fuzzy option set A on O is defined to be a set of ordered pairs

$$A = \{(x, \mu_A(x)) \mid x \in O\} \tag{1}$$

where $\mu_A(x)$ is called the membership function for the fuzzy option set A. The membership function maps each element in O to a membership value between 0 and 1

$$\mu_A : x \to \mu_A(x) \in [0, 1]. \tag{2}$$

In general, an option set O is composed of some options with discrete values. This can be clarified by the following example. If an individual has several attributes and each attribute can be specified to one of the five discrete options, i.e., 1, 3, 5, 7, 10. The fuzzy option set A with respect to "the belonging of 8" can be described as follows:

$$A = \{(1, 0.1), (3, 0.15), (5, 0.2), (7, 0.7), (10, 0.5)\}. \tag{3}$$

We can see that A is discrete and contains nonordered objects. The above example illustrates that the construction of a fuzzy option set depends on two factors: 1) the identification of a suitable option set and 2) an appropriate membership function. Basically, the selection of a membership function is subjective and derived from empirical analysis of the fuzzy data.

B. Option Entropy

Based on Shannon's entropy, which was first proposed as a measure of the uncertainty of random variables, we introduce opinion entropy in the following.

Let U be the collection of samples, O be the set of opinions, C be the set of conditional attributes, and D be the set of decisional attributes. $A \subseteq C$ is a subset of condition attributes. As shown in Table I, for this sample, at least two attributes may select the option "not at all." Thus, we use $n_\sigma(x)$ to indicate the number of sample x' attributes which choose option σ. An equivalence relation R_σ can be induced over O according to $n_\sigma(x)$

$$R_\sigma = \left\{\left(x_i, x_j\right) | \forall \sigma \in O, n_\sigma(x_i) = n_\sigma\left(x_j\right)\right\}. \tag{4}$$

Then, the partition U/R_σ will generate a set of equivalence classes $U_1, U_2, \ldots, U_n$, where the elements in $U_i (i =$

$1, 2, \ldots, n$) are difficult to distinguish because each class chooses the same number of σ.

Definition 2 (Option Entropy): Assuming $X_i (i = 1...n)$ is a set of samples which choose option σ for i attributes, the probability $p(X_i)$ with respect to X_i is calculated by $|X_i|/|U|$. Similarly, $Y_j (j = 1...m)$ is a set of samples which choose option φ for j attributes and the probability $p(Y_j)$ with respect to Y_j is calculated as $|Y_j|/|U|$. Then option entropy, joint entropy, and conditional entropy are defined as follows.

1) $H(\sigma) = -\sum_{i=1}^{n} p(X_i)\log p(X_i)$.
2) $H(\sigma, \varphi) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(X_i \cap Y_j)\log p(X_i \cap Y_j)$.
3) $H(\sigma|\varphi) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(X_i \cap Y_j)\log p(X_i|Y_j)$
$= -\sum_{i=1}^{n}\sum_{j=1}^{m} \frac{|X_i \cap Y_j|}{|U|}\log\frac{|X_i \cap Y_j|}{|Y_j|}$.

where $|X_i|$ denotes the cardinality of X_i.

Supposing $U_{\sigma(i)}$ is the subset of U, for each sample in this subset, there are i attributes choosing option σ. The subset $U^d_{\sigma(i)}$ represents the sample set in $U_{\sigma(i)}$ having the decision attribute d. The entropy of option σ on U is defined as follows:

$$E(\sigma, U) = -\sum_{i=1}^{|C|}\frac{|U_{\sigma(i)}|}{|U|}\sum_{d=1}^{|D|}\left(\frac{|U^d_{\sigma(i)}|}{|U_{\sigma(i)}|}\log_2\frac{|U^d_{\sigma(i)}|}{|U_{\sigma(i)}|}\right). \tag{5}$$

Theorem 1: Given a set of samples U with condition attributes C, and $d (d = 1, \ldots, |D|)$ is the decision attribute, then $H(d|\sigma) = E(\sigma, U)$. The proof is given as follows:

Proof:

$$\begin{aligned}
H(d|\sigma) &= -\sum_{i=1}^{|C|}\sum_{d=1}^{|D|} p\big(U_d \cap U_{\sigma(i)}\big)\log_2 p\big(U_d|U_{\sigma(i)}\big) \\
&= -\sum_{i=1}^{|C|}\sum_{d=1}^{|D|}\frac{|U_d \cap U_{\sigma(i)}|}{|U|}\log_2\frac{|U_d \cap U_{\sigma(i)}|}{|U_{\sigma(i)}|} \\
&= -\sum_{i=1}^{|C|}\frac{|U_{\sigma(i)}|}{|U|}\sum_{d=1}^{|D|}\frac{|U_d \cap U_{\sigma(i)}|}{|U_{\sigma(i)}|}\log_2\frac{|U_d \cap U_{\sigma(i)}|}{|U_{\sigma(i)}|} \\
&= -\sum_{i=1}^{|C|}\frac{|U_{\sigma(i)}|}{|U|}\sum_{d=1}^{|D|}\left(\frac{|U^d_{\sigma(i)}|}{|U_{\sigma(i)}|}\log_2\frac{|U^d_{\sigma(i)}|}{|U_{\sigma(i)}|}\right) \\
&= E(\sigma, U). \qquad \blacksquare
\end{aligned}$$

Definition 3 (Option Quality): Given a sample $x \in U_d$ and a subset $U^{\sim}_d = U - U_d$, $\text{count}(\sigma, U^{\sim}_d)$ is the number of samples in $U^{\sim}_d$ that are distinguished from the sample x by σ, i.e., the number of samples which belong to different classes from d and different values from x on σ. The average value of $\text{count}(\sigma, U^{\sim}_d)$ is computed by the following equation:

$$\text{aver}(\sigma, U^{\sim}_d) = \frac{1}{|D|}\sum_{i=1}^{|D|}\text{count}\big(\sigma, U^{\sim}_d\big). \tag{6}$$

Referring to attribute quality [25], we define the quality of option σ as follows:

$$Q(\sigma) = \begin{cases} +\infty & \text{aver}(\sigma, U^{\sim}_d) = 0 \\ \frac{E(\sigma,U)}{\text{aver}(\sigma,U^{\sim}_d)} * \text{SI}(\sigma, U) & \text{otherwise} \end{cases} \tag{7}$$

where $\text{SI}(\sigma, U)$ is computed by the following equation:

$$\text{SI}(\sigma, U) = -\sum_{i=1}^{n}\frac{|U_{\sigma(i)}|}{|U|}\log_2\frac{|U_{\sigma(i)}|}{|U|}. \tag{8}$$

Similar to the C4.5 algorithm [26], the term $\text{SI}(\sigma, U)$ is called the split information that is used to overcome the bias which the terms $E(\sigma, U)$ and $\text{count}(\sigma, U^{\sim}_d)$ have. It is used to correct the attributes with a lot of values.

$Q(\sigma)$ shows the amount of information that option σ contains for making a decision. The larger the value of $Q(\sigma)$ is, the more information σ contains for making a decision.

Theorem 2: Given a set of samples U, each attribute can choose a value from $1, \ldots, |O|$, where O represents the option set, then $H(\sigma) = \text{SI}(\sigma, U)$. The proof is given as follows:

Proof:

$$\begin{aligned}
H(\sigma) &= -\sum_{i=1}^{n} p(X_i)\log p(X_i) = -\sum_{i=1}^{n}\frac{|X_i|}{|U|}\log_2\frac{|X_i|}{|U|} \\
&= -\sum_{i=1}^{n}\frac{|U_{\sigma(i)}|}{|U|}\log_2\frac{|U_{\sigma(i)}|}{|U|} = \text{SI}(\sigma, U). \qquad \blacksquare
\end{aligned}$$

Definition 4 (Normalized Membership Function): The membership function is supposed to follow the Gaussian distribution. The maximal value of this function for a sample depends on the number of the central option which has been chosen. $\varphi(a, b)$ is employed to indicate whether these two variables are equal

$$\varphi(a, b) = \begin{cases} 1 & a = b \\ 0 & a \neq b. \end{cases} \tag{9}$$

Then, $\varphi(v, a)$ represents the number of elements whose values equal to a in the vector v

$$\varphi(v, a) = \sum_{i}\varphi(v_i, a) \tag{10}$$

where v_i represents the ith element in vector v. If $g^k_u(x)$ is the distribution function with the centrality u for the sample k, u is calculated by the following equation:

$$g^k_u(u) = \varphi(V_k, u) * Q(u) \tag{11}$$

where V_k is the kth vector with all attribute values. In addition, the item $Q(u)$ is involved, because it shows the effect of option u on making decisions. It is normalized by the following equation:

$$\text{norm}\Big(g^k_u(u)\Big) = \frac{g^k_u(u)}{\max_u\big(g^k_u(u)\big)} \in [0, 1]. \tag{12}$$

The normalized value of x with respect to the center u is

$$\begin{aligned}
f^k_u(x) &= \text{norm}\Big(g^k_u(x)\Big) \\
&= \frac{g^k_u(x)}{\max_u\big(g^k_u(u)\big)}.
\end{aligned} \tag{13}$$

This parameter shows the influence of option u on option x with respect to the sample k, i.e., the membership of option u on the point x. Fig. 1 shows the two membership functions

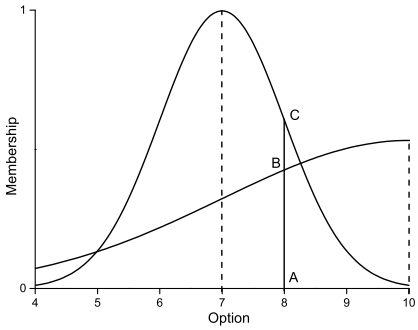


Fig. 1. Two membership functions with the options 7 and 10 as centrality, respectively.

with the option 7 and the option 10 as centrality, respectively. Option 7 has a larger membership value (the length of $\overline{AC}$) than option 10 (the length of $\overline{AB}$) at the option 8.

C. Influence Relation

Definition 5 (Interfering Degree): We use the concept of interfering degree to represent the interference between two options

$$r(\sigma_i, \sigma_j) = \frac{|n(\sigma_i) - n(\sigma_j)|}{\max(n(\sigma_i), n(\sigma_j))} \in [0, 1] \tag{14}$$

where $n(\sigma)$ represents the number of option σ. If there is a big difference between $n(\sigma_i)$ and $n(\sigma_j)$, the value $r(\sigma_i, \sigma_j)$ is close to 1; otherwise, $r(\sigma_i, \sigma_j)$ is approximately equal to 0. If two options are analogous and difficult to be distinguished, one of the options will be chosen frequently and the other one is not. In this situation, these two options can strongly influence each other.

Definition 6 (Influence Relation): Suppose W is a fuzzy option set on $O \times V$, where O and V represent different option sets, respectively. $O \times V \triangleq \{o \in O, v \in V\}$ is a Cartesian product. Its membership function is defined as follows:

$$W : O \times V \longrightarrow [0, 1] \tag{15}$$

$$(o, v) \longrightarrow W(o, v). \tag{16}$$

The membership function uncovers the influence relation between o in O and v in V, and is represented by $O \xrightarrow{W} V$. Actually, the relation between two options in an option set is a binary relation from O to O and is represented by $W(O \times O)$.

Definition 7 (Influence Relation Matrix): The influence matrix W is introduced to indicate the influence relation between any two options in the option set with respect to the sample k

$$W^k = \left[w_{ij}^k\right]_{n\times n}, \quad w_{ij}^k \in [0, 1], i \neq j \tag{17}$$

where w_{ij}^k is defined to be the directed influence relation from option σ_i to σ_j, and the value of w_{ij}^k is defined as follows:

$$w_{ij}^k = f_{\sigma_i}^k(\sigma_j) * r(\sigma_i, \sigma_j) \tag{18}$$

where $f_{\sigma_i}^k(\sigma_j)$ and $r(\sigma_i, \sigma_j)$ indicates the influence of option σ_i on σ_j for a specific sample and for the whole sample set, respectively.

Example 1: The relations among five options can be illustrated by the following influence relation matrix. The superscript k is omitted for simplicity

$$W = \begin{bmatrix} 1 & 0.65 & 0.3 & 0.1 & 0.05 \\ 0.6 & 1 & 0.5 & 0.15 & 0.1 \\ 0.7 & 0.8 & 1 & 0.6 & 0.3 \\ 0 & 0 & 0 & 1 & 0 \\ 0.01 & 0.05 & 0.15 & 0.28 & 1 \end{bmatrix}$$

where an element $w_{ij} = W(\sigma_i, \sigma_j)$ means the influence degree of σ_i on σ_j, and a jth column indicates a fuzzy option set with respect to "the belonging of option j." It is noticed that all elements of the fourth row are zero except $w_{4,4}$, which implies no one chooses option 4 for any attribute, and option 4 will not influence other options.

D. Fuzzy Attribute Reduction Based on Attribute Occurrence

To the best of our knowledge, the mental disease is widespread around the world, but the mental patients still account for a small proportion of the whole population. Reducing attributes is viewed as a difficult task due to the limit number of definite diagnoses.

An information system, as a basic concept in rough set theory, provides a convenient framework for the representation of objects in terms of their attribute values. An information system is a quadruple $S = (U, A, V, f)$, where U is a finite nonempty set of objects and is called the universe of discourse and A is a finite nonempty set of attributes, $V = \bigcup_{a\in A} V_a$ with V_a being the domain of a, and $f : U \times A \to V$ is an information function with $f(x, a) \in V_a$ for each $a \in A$ and $x \in U$. The system S can often be simplified as $S = (U, A)$.

For the information system $S = \{U, A\}$, suppose there is a set such that $B \subseteq A$. If: 1) $U/B = U/A$ and 2) $\forall a \in B, U/(B - \{a\}) \neq U/B$, then B is called a reduction set of S. There are usually multiple reduction sets in a given information system, and the intersection of all reduction sets is called the core set.

Given a decision table $S = \{U, C \cup D\}$, suppose there is a set such that $B \subseteq C$. If: 1) $\mathrm{POS}_B(D) = \mathrm{POS}_C(D)$ and 2) $\forall a \in B, \mathrm{POS}_{B-\{a\}}(D) \neq \mathrm{POS}_B(D)$, where POS represents the positive region, then B is a relative reduction set of S, and the intersection of all relative reduction sets is called relative core set [27].

According to the above two definitions, the first condition guarantees that the reduction set has the same distinguishing capability as the whole attribute set, and the second condition guarantees that there are no redundant attributes in a reduction set. A reduction set is called an exact reduction set if it satisfies the above two constraints, otherwise, it is only an approximate reduction set.

Definition 8 (Positive and Negative Decision Attribute Set): For an information system $S = \{U, C \cup D, V, f\}$ with the binary decision attribute, the positive decision attribute set, and the negative decision attribute set are represented by D_y and D_x, respectively, such that $D = D_y \cup D_x$.

In the clinical information system, D_y and D_x represent the definite diagnosis result without a disease and with a disease, respectively.

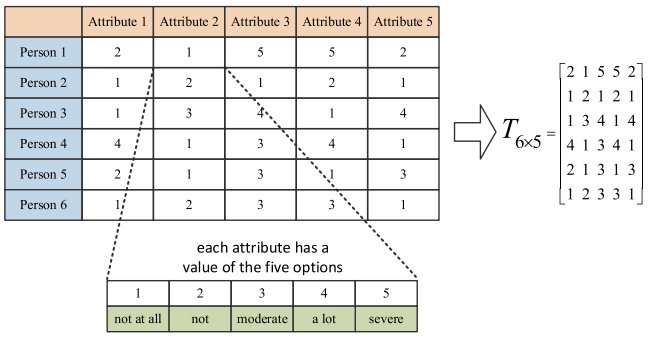


Fig. 2. Psychometry can be represented by a $n \times m$ matrix.

Definition 9 (Positive and Negative Object Set): According to D_y and D_x, The universe of discourse U is partitioned into two sets: U_y and U_x. If $|U_y| \gg |U_x|$, U_y is also divided into K sets: from U_{y1} to U_{yk}, such that $U_y = \sum_{i=1}^{k} U_{yi}$.

Proposition 1: Let $S = \{U, C \cup D\}$ be a decision table and $\mathrm{rem}(x)$ be the removed attribute set of x. If $U = (\cup_{i=1}^{K} U_{yi}) \cup U_x$, it follows.

1) $\mathrm{rem}(U) \subseteq (\cup_{i=1}^{K} \mathrm{rem}(U_{yi})) \cup \mathrm{rem}(U_x)$.
2) $\mathrm{rem}(U) \subseteq (\cup_{i=1}^{K} \mathrm{rem}(U_{yi} \cup U_x))$.

The proof is given as follows:

Proof:

1) Assuming $U_x \subseteq U, U_{yi} \subseteq U$
then $\mathrm{rem}(U) \subseteq \mathrm{rem}(U_x), \mathrm{rem}(U) \subseteq \mathrm{rem}(U_{yi})$.
Let $M = (\cup_{i=1}^{K} \mathrm{rem}(U_{yi})) \cup \mathrm{rem}(U_x)$
and then $\mathrm{rem}(U_x) \subseteq M, \mathrm{rem}(U_{yi}) \subseteq M$
finally we have $\mathrm{rem}(U) \subseteq M$.
2) Assuming $U_x \subseteq U, U_{yi} \subseteq U$
Let $P_i = U_x \cup U_{yi}$
then $\mathrm{rem}(U) \subseteq \mathrm{rem}(P_i)$.
Let $M = \cup_{i=1}^{K} \mathrm{rem}(U_{yi} \cup U_x)$,
and then $\mathrm{rem}(P_i) \subseteq M$
finally we have $\mathrm{rem}(U) \subseteq \mathrm{rem}(P_i) \subseteq M$. ■

IV. Hybrid Attribute Discrimination Method for Psychometric Data With Fuzziness

A. Framework Introduction

Before introducing our fuzzy method for key attribute discrimination, we first present the approach to formalize a psychometry. As shown in Fig. 2, a psychometry can be represented by a $n \times m$ matrix T. $T[i]$ represents the vector containing all the attributes with respect to a specific participant i.

In the clinical practice, numerical psychometries are used to assist doctors in making a definite diagnosis. A new method called ADPDF is proposed to extract key attributes from large-scale psychometric data. As shown in Fig. 3, the approach is partitioned into three steps.

1) The dataset is extracted from the psychometric database and the noisy and erroneous data are removed. The dataset consists of many valid participants, each of which contains some attributes. The participants select one from several options to indicate the degree of a somatic symptom.
2) Primary options are retained and other options are removed by Algorithm 1. We combine the removed options to the adjacent options. Then we reduce the attributes according to Algorithm 3.
3) Typical attribute discrimination methods are used to obtain key attributes. In order to sort the attributes, the linear regression should be chosen.

We will describe the main steps in detail in the following sections.

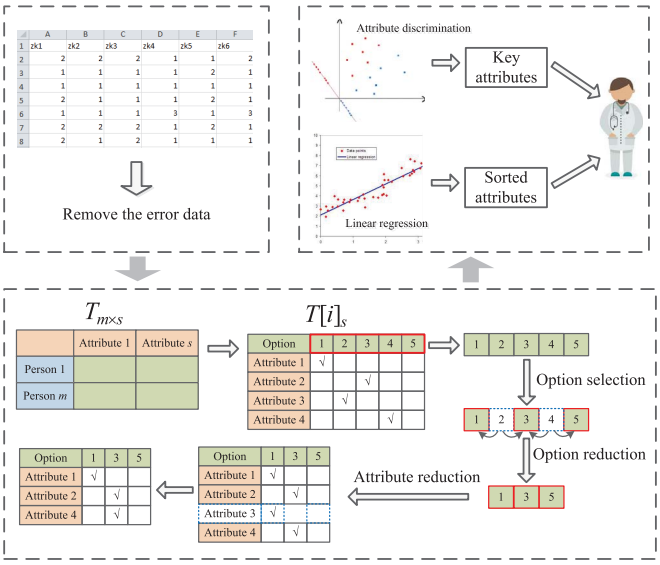


Fig. 3. Working mechanism of discriminating the key attributes based on the ADPDF method. It is divided into three parts, each of which also contains several steps.

B. Option Selection

All the options in a psychometry represent different degrees of attributes. The reserved options with almost the same intervals are chosen based on the diagnostic requirements. The option selection algorithm is shown in Algorithm 1.

The working mechanism of Algorithm 1 is given as follows.

1) In order to ensure the effectiveness of option reduction, the number of reserved options should be specified between 2 and $\lfloor (s+1)/2 \rfloor$. For example, there are 10 options, i.e., s = 10, ranging from 1 to 10. Table III shows a serial number of reserved options according to the total number of reserved options (lines 1–3).
2) Obtain the interval between two reserved options (line 4).
3) Add all reserved options except option s to the set O' (lines 5–7) because the interval between the last two options may be larger than *interval*.
4) Add the last option s to the set O' (line 8).

C. Fuzzy Option Reduction

Fuzzy option reduction method is used to reduce the fuzziness of options in a psychometry. It combines a fuzzy option to the reserved adjacent options. Any two adjacent options

Algorithm 1 Option Selection

Input: s: The initial number of options
s': The number of reserved options
Output: O': The set of reserved options

1: **if** $s' < 2$ or $s' > \lfloor (s+1)/2 \rfloor$ **then**
2: return;
3: **end if**
4: *interval*=$\lceil s/s' \rceil$;
5: **for** each $i \in [0, s'-2]$ **do**
6: add $1 + i * interval$ to the set O';
7: **end for**
8: add s to the set O'.

TABLE III
RELATION BETWEEN THE TOTAL NUMBERS AND THE SERIAL NUMBERS OF RESERVED OPTIONS

The total number	The serial number
5	1,3,5,7,10
4	1,4,7,10
3	1,5,10
2	1,10

share the maximal similarity with each other. The advantage is that the result of fuzzy option reduction approach can reflect the participants' subjective views and feelings, and reduce the probability of errors by randomly selecting fuzzy options.

After the reserved options are confirmed, the removed options should be combined to its adjacent reserved options. We use O and O' to represent the initial option set and reserved option set, respectively. Thus, $O'' = O - O'$ is the set of removed options.

If the option $\sigma \in O''$, the reserved adjacent options of σ are represented by σ_L and σ_R. The influence relations from σ_L to σ and from σ_R to σ are represented by $w(\sigma_L, \sigma)$ and $w(\sigma_R, \sigma)$, respectively. Then, the proportion of σ converted to σ_L is calculated by the following equation:

$$\mu = \frac{w(\sigma_L, \sigma)}{w(\sigma_L, \sigma) + w(\sigma_R, \sigma)}. \tag{19}$$

The number of the option σ chosen by participant i is obtained by the following equation:

$$n_A = \varphi(T[i], \sigma) \tag{20}$$

where $T[i]$ is the option vector of user i.

Then the numbers of the option σ converted to σ_L and σ_R are calculated by the following equation:

$$n_L = \begin{cases} \lceil n_A * \mu \rceil & w(\sigma_L, \sigma) \geqslant w(\sigma_R, \sigma) \\ \lfloor n_A * \mu \rfloor & w(\sigma_L, \sigma) < w(\sigma_R, \sigma) \end{cases} \tag{21}$$

$$n_R = n_A - n_L. \tag{22}$$

The ceil and floor symbols are used to ensure the option σ is probably changed to the option with larger influence. The option reduction algorithm is given in Algorithm 2.

The working mechanism of Algorithm 2 is given as follows.

1) O'' is the complement set of O' and is used to store the removed options (line 1).

Algorithm 2 Option Reduction

Input: O: The initial option set
O': The set of reserved options
T: The psychometric matrix ($m \times s$)
Output: T': The converted psychometric matrix ($m \times s$)

1: $O'' \leftarrow O - O'$;
2: **for** each $i \in [1, m]$ **do**
3: **for** each $\sigma \in O''$ **do**
4: $\sigma_L, \sigma_R \leftarrow$Neighbor($\sigma$);
5: $w(\sigma_L, \sigma) \leftarrow$CalcInfluence($\sigma_L, \sigma$);
6: $w(\sigma_R, \sigma) \leftarrow$CalcInfluence($\sigma_R, \sigma$);
7: $\mu \leftarrow w(\sigma_L, \sigma)/(w(\sigma_L, \sigma) + w(\sigma_R, \sigma))$;
8: $n_A \leftarrow \varphi(T[i], \sigma)$;
9: **if** $w(\sigma_L, \sigma) \geqslant w(\sigma_R, \sigma)$ **then**
10: $n_L \leftarrow \lceil n_A * \mu \rceil$;
11: **else**
12: $n_L \leftarrow \lfloor n_A * \mu \rfloor$;
13: **end if**
14: $n_R \leftarrow n_A - n_L$;
15: LabelL($T[i], \sigma, n_L$);
16: LabelR($T[i], \sigma, n_R$);
17: **end for**
18: **end for**
19: **for** each $i \in [1, m]$ **do**
20: **for** each $j \in [1, s]$ **do**
21: **if** $T[i, j]$ labeled as *ChangeL* **then**
22: $T'[i, j] \leftarrow \sigma_L$;
23: **else if** $T[i, j]$ labeled as *ChangeR* **then**
24: $T'[i, j] \leftarrow \sigma_R$.
25: **end if**
26: **end for**
27: **end for**

2) For a participant, label the specific options to be removed. These options will be combined to its adjacent options (lines 2–18).
3) Obtain the adjacent options σ_L and σ_R of σ (line 4).
4) Calculate the influence degrees from σ_L and σ_R to σ, respectively, according to (18). Here, only two nearest neighbors on both sides are taken into consideration because the membership function is supposed to follow the Gaussian distribution. The influence drops drastically when the distance grows (lines 5 and 6).
5) Calculate the proportion of σ converted to σ_L (line 7).
6) Calculate the numbers of σ converted to σ_L and σ_R, respectively, (lines 8–14).
7) Label the first n_L attributes with the option σ in $T[i]$ as *ChangeL* (line 15).
8) Label the rest of attributes with the option σ in $T[i]$ (the number is n_R) as *ChangeR* (line 16).
9) Convert the labeled attributes to the reserved options (lines 19–27).

D. K-Fold Attribute Reduction

The core attributes are very important in each subset. In general, it is difficult to remove abundant attributes due to

Algorithm 3 Attribute Reduction

Input: T: The psychometric matrix ($m \times s$)
K: The number of subsets
η: The number of removed attributes
Output: rem_{out}: The set of removed attributes

1: define array *rem*;
2: $X \leftarrow$NegSet(T);
3: $Y \leftarrow$PosSet(T);
4: $r \leftarrow$Column(Y)/K;
5: **for** each $i \in [1, K]$ **do**
6: $Y' \leftarrow$ NextRows(Y, r);
7: $T' \leftarrow Y' \cup X$;
8: *removed_attr* $\leftarrow$ GetRemoveAttr(T');
9: update *rem* by *removed_attr*;
10: **end for**
11: $rem_{out} \leftarrow$ the largest η elements in *rem*.

the interference of unnecessary participants. In order to obtain more accurate attributes from a psychometry containing a lot of attributes, we propose a new method for reduction called K-fold attribute reduction method. It negotiates the reduction results of different subsets to reduce the noise and discover the core attributes more accurately. As mentioned in the previous sections, a psychometric dataset is viewed as an information system $I = (U, A)$, where U is a nonempty finite set of participants. The new method for attribute reduction is described in Algorithm 3.

The working mechanism of Algorithm 3 is given as follows.

1) Define the array rem of s elements to calculate the removing occurrences of each attribute (line 1).
2) Calculate the samples with the negative and positive decisions in T, respectively, (lines 2 and 3).
3) Calculate the number of positive samples in a subset (line 4).
4) Repeat the operation of attribute reduction on K subsets, respectively. (lines 5–10).
5) Obtain the next subset Y' in Y with r positive samples (line 6).
6) The set T' is combined by Y' and X (line 7).
7) The attributes to be removed in T' are calculated by attribute reduction (line 8).
8) Once an attribute appears in *removed_attr*, the corresponding element in rem is increased by 1 (line 9).
9) The η attributes appearing the most frequently are viewed as the attributes to be removed (line 11).

The proposed ADPDF method removes some of abundant options and attributes which are not important even useless from original data. One option represents one case, and different values with respect to the option represent different degrees. Particularly, some attributes may have similar meanings. Therefore, combining abundant options, i.e., abundant attribute values, to the adjacent options, can keep the validity of data. Removing abundant attributes can help find the suspected mental patients.

V. Experiments and Analysis

A. Dataset Description

In this paper, two psychometric datasets are utilized to validate the performance of the proposed ADPDF method: 1) the PTSD dataset and 2) the dataset of mental disorders. These datasets were approved by the Ethics Committee of the West China Hospital, Sichuan University. Both the datasets are used throughout the experiments, and we only use the PTSD dataset to show the characteristics of ADPDF.

1) PTSD Dataset: Both PTSD and somatic symptoms were evaluated among adult and adolescent survivors six months after the Baoxing earthquake in 2013 using the PHQ-15 questionnaire [9], [28]. The PHQ-15 [29] is a self-report questionnaire composed of 15 attributes and used to measure somatic symptoms. Participants report the degrees of their somatic symptoms on a five-option questionnaire including "not bothered at all" (1), "bothered a little" (2), "bothered moderately" (3), "bothered a lot" (4), and "bothered severely" (5). Actually, the two questions of the questionnaire pertaining to menstruation and sexuality were not taken into consideration. PHQ-13 questionnaire, a short version of the PHQ-15 questionnaire, is used. It contains the following attributes: stomach pain (1); back pain (2); pain in arms, legs, or joints (3); head pain (4); chest pain (5); dizziness (6); fainting spells (7); feeling heart pound/race (8); shortness of breath (9); bowel problems (10); nausea, gas, or indigestion (11); tired & low energy (12); and trouble sleeping pound/race (13). The Chinese version of the PHQ-13 has demonstrated a satisfactory level of internal consistency and reliability in the general population of China [30]. Table IV demonstrates the data form of the PTSD dataset. A participant chooses a value between 1 and 5 for each attribute, and finally the psychiatrist makes a diagnostic decision which is represented by 0 (not have) or 1 (have). For example, the first participant chooses option 1 for the attribute "stomach pain," which means stomach pain has not bothered him/her at all. Similarly, he/she chooses option 4 for the attribute "chest pain," which indicates chest pain has bothered him/her with a larger intensity. Based on the collected data, a psychiatrist can make a final diagnostic decision that the participant owns no disease of PTSD.

After removing abundant data, the dataset contains 3099 participants. 51.9% were female, 87.5% were Han Chinese and 12.5% were Chinese minorities. Their ages ranged from 14 to 91 years. The correlation coefficients between different somatic symptoms and probable PTSD are shown in Fig. 4. The trouble sleeping, feeling tired, and nausea, gas or indigestion, are viewed as three primary somatic symptoms of PTSD due to their high correlations with PTSD.

2) Dataset of Mental Disorders: A large-scale epidemiological surveys of mental disorders in children and adolescents were conducted in Sichuan. This paper is one of the China's first nationwide epidemiological projects [31], which aims to understand the incidence of behavioral problems and find the risk factors in psychiatric disorders in students aged from 6 to 16 years in Sichuan province.

There are 20 752 students surveyed by answering the CBCL questionnaire (Achenbach's child behavioral checklist). Then

TABLE IV
DATA FORM OF THE PTSD DATASET

Participant No.	Stomach pain	Back pain	Pain in arms, legs	Head pain	Chest pain	...	Trouble sleeping	Diagnosis
1	1	1	2	1	4	...	1	0
2	5	1	5	2	4	...	2	1
...	...	...	...	...	...	...	...	...
3099	1	1	2	1	3	...	3	0

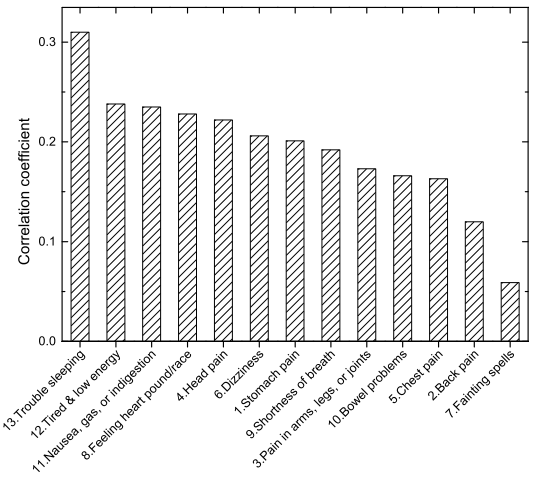


Fig. 4. Correlation coefficients between different somatic symptoms and probable PTSD.

the probably disordered were given definite diagnosis by the interviews of psychiatrists.

The standardized Chinese version of CBCL was used with 113 questions selected from the behavioral problems section. Participants rate the child's behavior on a 3-point questionnaire including "not true" (0), "somewhat or sometimes true" (1), and "very true or often true" (2), and are instructed to rate the behavior as it occurs within the previous two months.

Instead of analyzing all the categories of mental disorders, we extract the questions about the aggressive behavior from the total 113 questions.

The diagnosis process based on the two categories of psychometries can be viewed as an information system $S = \{U, C \cup D, V, f\}$ with the binary decision attribute D. The decision value 0, i.e., $d = 0$, indicates the person does not have the disease, otherwise, $d = 1$ indicates the person has the disease authentically. The condition attribute C consists of the values from 1(stomach pain) to 13(trouble sleeping pound/race) for the PTSD dataset and the values from 1 to 23 for the dataset of mental disorders. $V = \bigcup_{a \in A} V_a$ is composed of V_a, the domain of a, which is the set of the integral option values from 1 to 5 and from 1 to 3, respectively. Table V shows the brief information of these two datasets.

B. Characteristics of Datasets

Due to the similar characteristics of the two datasets, we only conducted experiments on the PTSD dataset in the following sections. It is worthwhile to notice that most of the participants tend to select one to three options, instead of selecting all the five options on average.

TABLE V
DETAILED INFORMATION OF THE TWO DATASETS

	PHQ-13	CBCL
Target disease	PTSD	Aggressive behaviors
Number of samples	3099	20752
Number of attributes	13	23
Number of options	5	3

TABLE VI
OPTION COMBINATIONS OF THE LARGEST OCCURRENCE FREQUENCIES. THESE FOUR COLUMNS ARE NUMBER OF OPTIONS, TOTAL OCCURRENCE PERCENTAGE, OPTION COMBINATION, AND EACH OCCURRENCE PERCENTAGE

Number of opt.	Total occ. pct.	Opt. comb.	Each occ. pct.
1 option	38.0%	1	38.0%
2 options	35.5%	1&2	31.0%
		1&4	4.5%
3 options	20.6%	1&2&3	13.2%
		1&2&4	3.6%
		1&3&4	2.7%
		1&3&5	1.1%

TABLE VII
INITIAL CORRELATION COEFFICIENTS BETWEEN DIFFERENT OPTIONS

	option 1	option 2	option 3	option 4	option 5
option 1	1				
option 2	-0.761	1			
option 3	-0.508	0.081	1		
option 4	-0.359	-0.009	0.378	1	
option 5	-0.208	-0.031	0.124	0.270	1

Table VI shows the option combinations of the largest occurrence frequencies. It implies that too many options are not helpful to a participant. The participant only needs two or three alternative options to represent the degrees of symptoms. Tables VII and VIII show the correlation coefficients of any two options before and after the option reduction.

As shown in Table VII, an option has a higher correlation with its neighbors than far ones. For example, option 2 has a higher correlation with option 3 than option 4 or 5. However, option 1 is an exception because it represents the normal state and is usually treated as the default option. When a person chooses option 1 for many attributes, he tends to ignore some mild symptoms, i.e., he is less likely to choose options close

TABLE VIII
FINAL CORRELATION COEFFICIENTS BETWEEN DIFFERENT OPTIONS AFTER OPTION REDUCTION

	option 1	option 3	option 5
option 1	1		
option 3	-0.737	1	
option 5	-0.563	-0.279	1

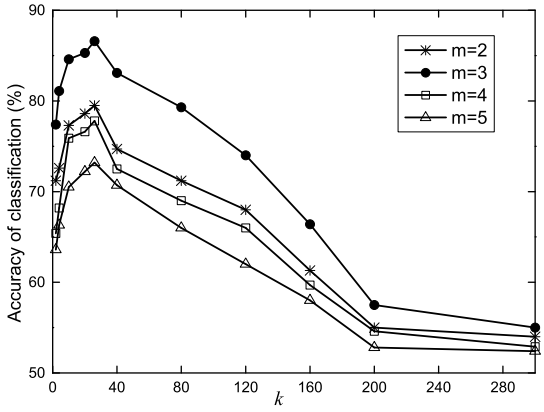


Fig. 5. Relation between the classification accuracy and two parameters including the number of reserved options in the option reduction and the number of divided subsets in the attribute reduction.

to option 1, such as option 2 or 3. Other options far from option 1, such as option 4 or 5 are usually slightly affected.

Reducing and merging options reduces the correlation between options. It is easier for a person to distinguish the options of small correlation coefficient according to Table VIII. Especially, the minimal coefficient less than 0 mean that the option is a clear choice with little fuzziness.

C. Parameter Setup

Core options and attributes are reserved to reduce the fuzziness of psychometric data. The number of reserved options (represented by m) in the option reduction and the number of partitioned subsets (represented by k) in the attribute reduction are treated to be essential parameters. Fig. 5 depicts the classification accuracy when specifying different m and k values. Tenfold cross validation was applied to evaluate the classification accuracy in the following study.

As shown in Fig. 5, the classification accuracy shows a rising tendency at the beginning, and then drops when k becomes larger. The classification accuracy reaches to the best performance when $k = 30$. This occurs as a result of about 3000 samples involved in the dataset, and we can conclude that the best size of a subset for attribute reduction is about 100 (that is 3000/30) when a sample has 13 attributes. In addition, when reserving three options ($m = 3$), i.e., option 1, 3, and 5, it has a better classification performance than reserving only two or four options, even than the initial state with five options. The accuracy is improved by 2.6%–14.8% when the initial 5 options are reduced to 3 options. The reason can be explained as follows: three options with the degrees of not bothered at all (1), bothered moderately (3), and bothered severely (5) in the PTSD dataset are the optimal option combination to help a person make a decision. Reserving two options may lose some important information. Meanwhile, the combination of five options is the original state without option reduction. It contains fuzzy information that may cause inaccurate decisions.

TABLE IX
CLASSIFICATION PERFORMANCE OF DIFFERENT k, m AND INITIAL OPTION DISTRIBUTIONS

Subset information				Metrics	Av.	SD	Max.	Min.
Av.	SD	K	m					
1.1	0.24	60	3	Precision(%)	**66.2**	11.3	**82.8**	50.3
				Recall(%)	63.3	**10.1**	77.6	50.6
				Accuracy(%)	64.7	10.2	78.1	**51.8**
1.1	0.24	30	3	Precision(%)	**67.7**	13.8	85.2	51.3
				Recall(%)	65.6	**7.7**	**85.3**	**52.9**
				Accuracy(%)	65.7	9.3	81.0	52.8
2.3	0.51	60	3	Precision(%)	**68.1**	13.6	**85.3**	51.5
				Recall(%)	67.1	8.9	81.5	52.5
				Accuracy(%)	65.4	**5.6**	74.2	**54.9**
2.3	0.51	30	3	Precision(%)	**70.5**	16.2	**88.3**	51.0
				Recall(%)	67.9	10.4	82.5	52.1
				Accuracy(%)	65.7	**7.9**	79.7	**54.2**
3.1	0.60	60	3	Precision(%)	**71.3**	15.9	88.1	51.1
				Recall(%)	69.1	7.7	**84.7**	54.2
				Accuracy(%)	66.2	**5.4**	75.1	**56.9**
3.1	0.60	30	3	Precision(%)	**73.3**	12.9	**88.1**	54.9
				Recall(%)	71.9	**8.8**	85.0	**57.2**
				Accuracy(%)	66.7	9.2	82.3	52.8
4.0	0.62	60	3	Precision(%)	**81.0**	9.5	**93.6**	68.2
				Recall(%)	78.8	**5.9**	87.5	**68.7**
				Accuracy(%)	78.1	7.4	89.0	67.1
4.0	0.62	30	3	Precision(%)	**85.8**	**6.2**	94.4	**75.6**
				Recall(%)	82.0	7.3	93.3	70.8
				Accuracy(%)	81.9	12.1	**96.2**	63.6

In order to evaluate the performance of classification algorithms, we use the following popular measurements: precision, recall, and accuracy [32]. The dataset is divided into observed positive and negative instances. In this paper, the positive class is defined as the normal people without a disease, and the negative class indicates the people with a disease.

The initial option distribution is another significant factor affecting the classification accuracy besides the two parameters of k and m. Table IX shows the classification performances by different combination of parameters. The bold font means the best performance of the compared terms. The subsets are specified under different k, m and distribution information including average values and standard deviations of initial options (i.e., the first four columns). Then, the classification performance metrics of each subset including precision, recall and accuracy are shown in the rest of the columns. To depict these three metrics clearly, we use other four parameters, i.e., average value, standard deviation, maximal and minimal values with respect to the corresponding measurements.

As shown in Table IX, a subset with a larger average value of options implies the samples in this subset are more

TABLE X
PERFORMANCES OF ALGORITHMS UNDER DIFFERENT LEVELS OF REDUCTION

	ADPDF	YONA	NOYA	NONA
option reducton	yes	yes	-	-
attribute reduction	yes	-	yes	-
removed options	2,4	2,4	-	-
removed attributes	1,10,12,13	-	1,13	-
primary attributes	7,8,9	1,8,13	8,9,12	1,8,13
time(option reduction)	**44s**	**44s**	-	-
time(attribute reduction)	**31min**	-	33min	-
prediction accuracy	**85.1%**	81.3%	83.2%	80.6%
training time	**23.7s**	24.0s	24.2s	25.3s

likely to suffer from PTSD. The subset contains more samples with PTSD and the number of samples with the negative and positive decisions are comparable. Thus, It shows a better classification performance than a subset with small average value of options. For a specific subset, if we partition the dataset into more subsets, the performance will become worse. The result is consistent with the dropping tendency shown in Fig. 5. In addition, The subset with a larger average option always has a larger standard deviation. The reason is that the people with PTSD will choose large options, such as options 4 and 5, more frequently than those without PTSD. However, no definite relation is discovered between the standard deviations of classification performance measurements with respect to different subsets.

Based on the analysis in this section, we set the experimental parameters with respect to the two datasets as follows.

1) *PTSD Dataset:* $m = 3$ and $k = 30$, where m represents the number of reserved options and k is the number of partitioned subsets.
2) *Dataset of Aggressive Behaviors:* $m = 2$ and $k = 30$.

D. Algorithm Evaluation Under Different Reduction Levels

To validate the proposed algorithm, we observe it under the following circumstances with the same parameters, $k = 30$ and $m = 3$.

1) Using the option reduction and the k-fold attribute reduction (ADPDF).
2) Using the option reduction, and the attribute reduction without k-fold attribute reduction (YONA for short).
3) Not using the option reduction, but using the k-fold attribute reduction (NOYA for short).
4) Not using the option reduction, but only using the attribute reduction without k-fold attribute reduction (NONA for short).

As we can see, the difference between these four methods is whether using the option reduction or the k-fold attribute reduction. The experimental results of these methods are shown in Table X.

As shown in Table X, the same options, i.e., options 2 and 4 are selected and removed for both ADPDF and YONA. The operation of option reduction changes the option's distribution and affect the following options, such as attribute reduction. NOYA and NONA have no operation of option reduction, so five options are all reserved. The fuzzy information makes it difficult to accurately remove abundant attributes and these two methods show worse classification performance than ADPDF and YONA.

Four attributes, i.e., trouble sleeping (13), feeling tired (12), stomach pain (1), and bowel problems (10), are viewed as abundant attributes and removed in the phase of attribute reduction. However, trouble sleeping (13) and feeling tired (12) of these four attributes to be removed are treated as primary factors for PTSD due to their high correlations with PTSD in Fig. 4. The reasons can be explained by the following reasons: these attributes are more likely to choose large options than other attributes in all cases. In other words, these attributes have little information for a decision. For PTSD, the symptoms, such as trouble sleeping are prevalent in people even without the earthquake experience. The earthquake may increase the proportion of trouble-sleeping people or make the existing trouble-sleeping more severe. In addition, the key attributes extracted by the linear regression and the classification accuracy are totally different among these four methods. ADPDF shows the best performance than other approaches. These findings are apparent because the options and attributes removed by these four methods are different, and ADPDF contains more valuable information.

The number of options and the number of attributes are usually limited, for example, s is 13 and n is five in the PHQ-13 questionnaire. When the two numbers are ignored, the computational complexity of the option reduction can be simplified to $\mathcal{O}(m^2)$, where m denotes the number of samples. In addition, the computational complexity of the attribute reduction is $\mathcal{O}(k * m^2)$, where k is the number of subsets. Due to the scope of k value, for example $k = 10$ in our experiments, the complexity is simplified to $\mathcal{O}(m^2)$. Therefore, the operation of attribute reduction and the operation of option reduction have the similar computational complexity.

All the four algorithms have similar classification accuracy. The option reduction and attribute reduction can remove some abundant information, and make a more accurate decision. However, option reduction can lead to the loss of information to some extent. By taking these two factors into consideration simultaneously, the methods with option reduction have higher accuracy than those without option reduction. Furthermore, the results in the last row indicate that all the algorithms have similar training time.

E. Comparison of Different Attribute Discrimination Methods

In our experiments, three classical attribute discrimination algorithms are used as baseline methods to verify the performance of the proposed ADPDF method.

1) *Principal Component Analysis (PCA) [33]:* This statistical method uses an orthogonal transformation to convert a set of possibly correlated variables into a set of values of linearly uncorrelated variables which are called principal components. The projection of data with the largest variance indicates the most principal components.

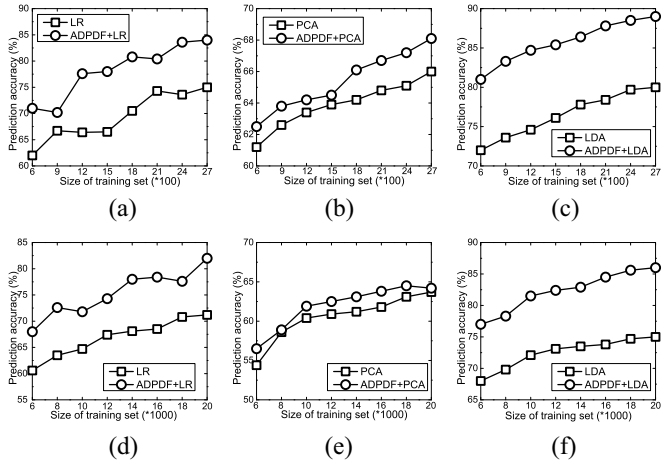


Fig. 6. Classification accuracy of three attribute discrimination algorithms under different sizes of training set. (a)–(c) Results using the PTSD dataset with five reserved attributes. (d)–(f) Results using the dataset of aggressive behaviors with ten reserved attributes. Each algorithm is compared under two cases: with ADPDF and without ADPDF.

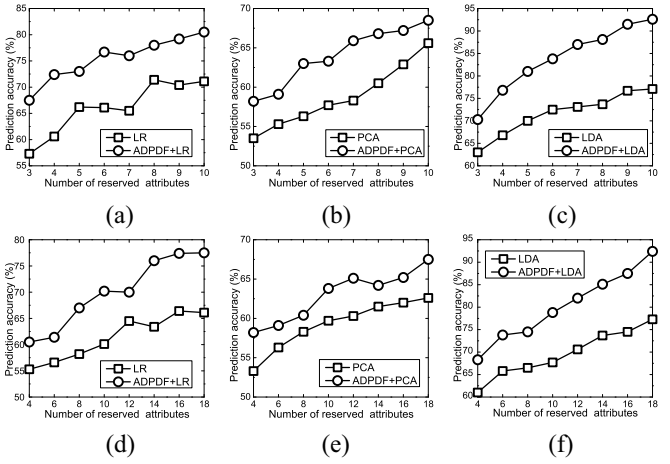


Fig. 7. Classification accuracy of three attribute discrimination algorithms under different numbers of reserved attributes. (a)–(c) Results using the PTSD dataset with 3099 samples. (d)–(f) Results using the dataset of aggressive behaviors with 20 752 samples. Each algorithm is compared under two cases: with ADPDF and without ADPDF.

2) *Linear Discriminant Analysis (LDA) [34]:* This method is used to find a linear combination of features that characterizes or separates two or more classes of objects.
3) *Fuzzy Discriminant Analysis (FDA) [35]:* This method is applied to the analysis of fuzzy data based on their degrees of importance. Overlapping data are assigned low membership values and can be easily separated.

The linear regression method was also involved in this set of experiments due to its capacity of sorting attributes. Each attribute with several options in a psychometry can be viewed as a feature with several probable values in the baseline methods. The task of extracting key attributes is similar to dimensionality reduction in attribute discrimination algorithms. Therefore, these methods can be compared directly in the same baseline.

The ADPDF method reserved three options from five original options and adapted 30 subsets to the attribute reduction with respect to the PTSD dataset. Two options were reserved from three original options and 210 subsets were used with respect to the dataset of aggressive behaviors. Figs. 6 and 7 show the comparison results of the three algorithms on these two datasets, respectively.

As shown apparently in Figs. 6 and 7, the classification accuracy increases by at least 3.5% due to ADPDF with respect to these three attribute discrimination algorithms except PCA. PCA retains as many features as possible and performs worse at classification. Figs. 6(b) and (e) and 7(b) and (e) show that ADPDF can hardly improve the classification performance of PCA. Although the operation of option reduction causes the loss of information, the combination of fuzzy options makes the option distribution clear. In addition, LDA is sensitive to the parameters especially the number of reserved attributes.

From Fig. 7(c) and (f), we can see that the number of reserved attributes improves the classification performance apparently with the gap from 7.3% at the least reserved attributes to 15.5% at the most reserved attributes with respect to the two datasets. Furthermore, Figs. 6(a) and (d) and 7(a) and (d) show the classification accuracy is increased by 3.5%–11.9%. Because the linear regression is not a conventional attribute discrimination algorithm, its performance fluctuates drastically.

However, ADPDF spends additional time besides the operation of classification. The efficiency of ADPDF is a little bit lower than other algorithms, because it needs to compare attributes one by one, which is costly. Other attribute reduction methods, such as discernibility matrix method [36] and information entropy method [37] are more efficient. In order to compare the attribute discrimination methods based on the proposed ADPDF to the typical fuzzy algorithm, i.e., FDA, we conduct experiments by reserving the same number of attributes to verify the efficacy of different algorithms specifically and clearly. Fig. 8 shows the comparison results for classification accuracy and time consumption on the two datasets, respectively.

As shown in Fig. 8(a) and (c), LDA with ADPDF has better classification performances than the other three fuzzy methods. In particular, PCA with ADPDF has a worse effect even than FDA because PCA can retain the most features in attribution discrimination but perform worse at classification. The gap between LDA with ADPDF and PCA with ADPDF changes from 12.1% to 24.1% with respect to the PTSD dataset and changes from 10.1% to 24.9% with respect to the dataset of aggressive behaviors. Although the linear regression with ADPDF is 7.8%–14.9% less accurate than LDA with ADPDF, it can sort all the attributes based on their importance besides outputting three or five reserved attributes. Moreover, reserving more attributes will increase the classification accuracy of all the fuzzy methods apparently.

The computational complexity of all these algorithms including the reduction and training processes are $\mathcal{O}(n^2)$. As depicted in Fig. 8(b) and (d), a method of high classification accuracy usually costs more time, and vice versa. The only exception is that LR with ADPDF has better classification and time performance than PCA with ADPDF since PCA is not presented for the classification purpose. Furthermore,

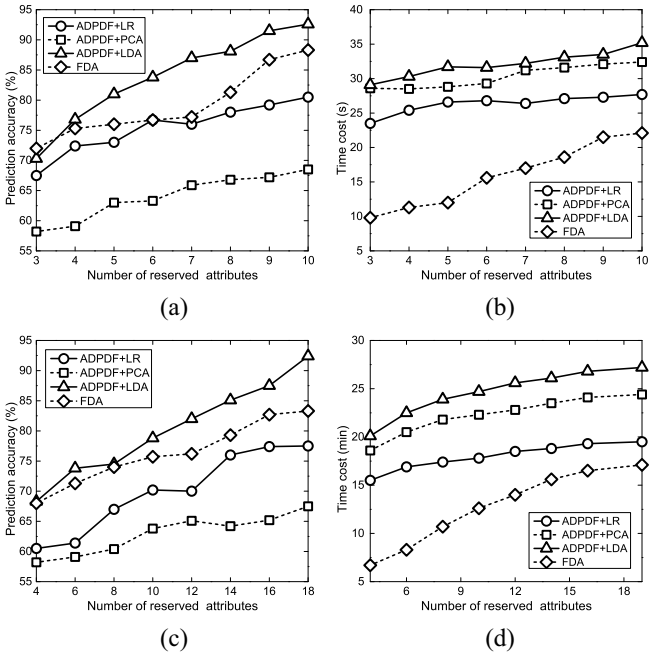


Fig. 8. Comparison results for accuracy and time of four fuzzy attribute discrimination methods. The dataset consists of 3099 samples and five attributes are reserved. (a) Comparison of the prediction accuracy (PTSD). (b) Comparison of the time cost (PTSD). (c) Comparison of the prediction accuracy (aggressive behaviors). (d) Comparison of the time cost (aggressive behaviors).

the three algorithms with ADPDF cost much more time than FDA with gaps larger than 5.6 s and 2.4 min on the two datasets, respectively. Because the reduction operations are time-consuming.

VI. Conclusion

In this paper, we present new concepts, i.e., option entropy and option influence degree, to describe the relation and distribution of options. In order to remove the abundant information of psychometric data, we propose a hybrid attribute discrimination method for psychometric data with fuzziness called ADPDF that plays an essential role in classification. The experimental results based on two clinical datasets show that ADPDF decreases the correlation between options effectively. Three reserved options and 100 samples per subset show the best classification performance. Finally, we compare typical attribute discrimination algorithms by using the proposed method. The comparison results reveal that the computational complexity of all these methods is similar and our method can effectively improve the classification performance.

The proposed ADPDF approach has some disadvantages that need to be handled in the future.

1) ADPDF provides a novel solution to extract the valuable information from psychometric data, however, fine-grained factors, such as age, gender, and living environment, *et al.*, are not taken into full consideration. Analyzing these effects by means of fuzzy sets and rough sets is expected to be a valuable and meaningful work.
2) The phase of selecting fuzzy options is very easy and may affect the prediction accuracy. In the future, we can use the Dempster–Shafer evidence theory to improve the accuracy of fuzzy option selection, which can help improve the accuracy of attribute discrimination.

References

[1] P. Dagar, A. Jatain, and D. Gaur, "Medical diagnosis system using fuzzy logic toolbox," in *Proc. Int. Conf. Comput. Commun. Autom.*, 2015, pp. 193–197.
[2] M. R. Sumathi and D. B. Pooma, "Prediction of mental health problems among children using machine learning techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 1, pp. 552–557, 2016.
[3] H. L. Chen *et al.*, "An efficient diagnosis system for detection of Parkinson's disease using fuzzy *k*-nearest neighbor approach," *Expert Syst. Appl.*, vol. 40, no. 1, pp. 263–271, 2013.
[4] C. S. Son, Y. N. Kim, H. S. Kim, H. S. Park, and M. S. Kim, "Decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches," *J. Biomed. Informat.*, vol. 45, no. 5, pp. 999–1008, 2012.
[5] E. E. Mcginty, J. Baller, S. T. Azrin, D. Juliano-Bult, and G. L. Daumit, "Quality of medical care for persons with serious mental illness: A comprehensive review," *Schizophrenia Res.*, vol. 165, nos. 2–3, pp. 227–235, 2015.
[6] R. C. Kessler *et al.*, "The global burden of mental disorders: An update from the who world mental health (WMH) surveys," *Epidemiol. Psychiatr. Sci.*, vol. 18, no. 1, pp. 23–33, 2009.
[7] E. Castarlenas, M. P. Jensen, C. L. von Baeyer, and J. Miró, "Psychometric properties of the numerical rating scale to assess self-reported pain intensity in children and adolescents: A systematic review," *Clin. J. Pain*, vol. 33, no. 4, pp. 376–383, 2017.
[8] S. Shannon, G. Breslin, B. Fitzpatrick, D. Hanna, and D. Brennan, "Testing the psychometric properties of kidscreen-27 with Irish children of low socio-economic status," *Qual. Life Res.*, vol. 26, no. 4, pp. 1081–1089, 2017.
[9] J. Zhang, S. Zhu, C. Du, and Y. Zhang, "Posttraumatic stress disorder and somatic symptoms among child and adolescent survivors following the Lushan earthquake in China: A six-month longitudinal study," *J. Psychosom. Res.*, vol. 79, no. 2, pp. 100–106, 2015.
[10] F. Dazzi, A. Shafer, and M. Lauriola, "Meta-analysis of the brief psychiatric rating scale—Expanded (BPRS-E) structure and arguments for a new version," *J. Psychiat. Res.*, vol. 81, pp. 140–151, Oct. 2016.
[11] J. López and S. Maldonado, "Group-penalized feature selection and robust twin SVM classification via second-order cone programming," *Neurocomputing*, vol. 235, pp. 112–121, Apr. 2017.
[12] W. S. Du and B. Q. Hu, "Attribute reduction in ordered decision tables via evidence theory," *Inf. Sci.*, vols. 364–365, pp. 91–110, Oct. 2016.
[13] Q. Hu, Z. Xie, and D. Yu, "Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation," *Pattern Recognit.*, vol. 40, no. 12, pp. 3509–3521, 2007.
[14] J. Symanzik, K. H. Choi, G. S. Mun, and J. Y. Ahn, "Sensitivity of an interview chart for medical diagnoses of primary headaches," *Int. J. Innov. Comput. Inf. Control*, vol. 8, no. 10, pp. 7133–7142, 2012.
[15] C.-Y. Wang and S.-M. Chen, "Multiple attribute decision making based on interval-valued intuitionistic fuzzy sets, linear programming methodology, and the extended TOPSIS method," *Inf. Sci.*, vols. 397–398, pp. 155–167, Aug. 2017.
[16] Z. Xu and M. Xia, "Distance and similarity measures for hesitant fuzzy sets," *Inf. Sci.*, vol. 181, no. 11, pp. 2128–2138, 2011.
[17] H. Choi, K. Han, K. Choi, and J. Ahn, "A fuzzy medical diagnosis based on quantiles of diagnostic measures," *J. Intell. Fuzzy Syst.*, vol. 31, no. 6, pp. 3197–3202, 2016.
[18] R. Y. Masri and H. M. Jani, "Employing artificial intelligence techniques in mental health diagnostic expert system," in *Proc. Int. Conf. Comput. Inf. Sci.*, 2012, pp. 495–499.
[19] R. M. Rahman and F. Afroz, "Comparison of various classification techniques using different data mining tools for diabetes diagnosis," *J. Softw. Eng. Appl.*, vol. 6, no. 3, pp. 85–97, 2013.
[20] F. L. Seixas, B. Zadrozny, J. Laks, A. Conci, and D. C. M. Saade, "A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimers disease and mild cognitive impairment," *Comput. Biol. Med.*, vol. 51, no. 7, pp. 140–158, 2014.
[21] A. Khemphila and V. Boonjing, "Parkinson's disease classification using neural network and feature selection," *World Acad. Sci. Eng. Technol.*, vol. 64, no. 4, pp. 15–18, 2012.

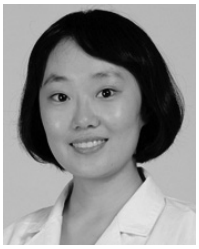
[22] F. Dabek and J. J. Caban, "A neural network based model for predicting psychological conditions," in *Proc. Int. Conf. Brain Informat. Health*, London, UK, 2015, pp. 252–261.
[23] V. Prasad, T. S. Rao, and M. S. P. Babu, "Thyroid disease diagnosis via hybrid architecture composing rough data sets theory and machine learning algorithms," *Soft Comput.*, vol. 20, no. 3, pp. 1179–1189, 2016.
[24] B. Saha, T. Nguyen, D. Phung, and S. Venkatesh, "A framework for classifying online mental health-related communities with an interest in depression," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 4, pp. 1008–1015, Jul. 2016.
[25] J. Wang, P. Zhang, G. Wen, and J. Wei, "Classifying categorical data by rule-based neighbors," in *Proc. IEEE 11th Int. Conf. Data Min.*, Vancouver, BC, Canada, 2011, pp. 1248–1253.
[26] C. J. Mantas, J. Abellán, and J. G. Castellano, "Analysis of credal-C4.5 for classification in noisy domains," *Expert Syst. Appl.*, vol. 61, pp. 314–326, Nov. 2016.
[27] F. Wang, J. Liang, and C. Dang, "Attribute reduction for dynamic data sets," *Appl. Soft Comput.*, vol. 13, no. 1, pp. 676–689, 2013.
[28] J. Zhang *et al.*, "Prevalence and risk factors of posttraumatic stress disorder among teachers 3 months after the Lushan earthquake: A cross-sectional study," *Medicine*, vol. 95, no. 29, 2016, Art. no. e4298.
[29] S.-C. Liao *et al.*, "The relation between the patient health questionnaire-15 and DSM somatic diagnoses," *BMC Psychiat.*, vol. 16, no. 1, p. 351, 2016.
[30] S. Lee, Y. L. Ma, and A. Tsang, "Psychometric properties of the Chinese 15-item patient health questionnaire in the general population of Hong Kong," *J. Psychosom. Res.*, vol. 71, no. 2, pp. 69–73, 2011.
[31] Y. Qu, H. Jiang, N. Zhang, D. Wang, and L. Guo, "Prevalence of mental disorders in 6-16-year-old students in Sichuan Province, China," *Int. J. Environ. Res. Public Health*, vol. 12, no. 5, pp. 5090–5107, 2015.
[32] J. F. Díez-Pastor, J. J. Rodríguez, C. García-Osorio, and L. I. Kuncheva, "Random balance: Ensembles of variable priors classifiers for imbalanced data," *Knowl. Based Syst.*, vol. 85, pp. 96–111, Sep. 2015.
[33] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, p. 11, 2009.
[34] A. Sharma and K. K. Paliwal, "A deterministic approach to regularized linear discriminant analysis," *Neurocomputing*, vol. 151, no. 1, pp. 207–214, 2015.
[35] Z.-P. Chen, J.-H. Jiang, Y. Li, Y.-Z. Liang, and R.-Q. Yu, "Fuzzy linear discriminant analysis for chemical data sets," *Chemometrics Intell. Lab. Syst.*, vol. 45, nos. 1–2, pp. 295–302, 1999.
[36] Y. Yao and Y. Zhao, "Discernibility matrix simplification for constructing attribute reducts," *Inf. Sci.*, vol. 179, no. 7, pp. 867–882, 2009.
[37] J.-H. Dai *et al.*, "Attribute reduction in interval-valued information systems based on information entropies," *Front. Inf. Technol. Electron. Eng.*, vol. 17, no. 9, pp. 919–928, 2016.

Xi Xiong received the B.S. and M.S. degrees in electronic engineering from the Beijing Institute of Technology, Beijing, China, and the Ph.D. degree in information security from Sichuan University, Chengdu, China, in 2013.

He is currently a Lecturer with the School of Cybersecurity, Chengdu University of Information Technology, Chengdu. He is currently involved in medical data mining. His current research interests include data mining, social computing, and machine learning.

Shaojie Qiao received the B.S. and Ph.D. degrees in computer science and technology from Sichuan University, Chengdu, China, in 2004 and 2009, respectively.

From 2007 to 2008, he was a Visiting Scholar with the School of Computing, National University of Singapore, Singapore. From 2009 to 2012, he was a Post-Doctoral Researcher with the Post-Doctoral Research Station, Engineering of Traffic Transportation, Southwest Jiaotong University, Chengdu. He is currently a Professor with the School of Cybersecurity, Chengdu University of Information Technology, Chengdu. He has led several research projects in the areas of databases and data mining. He has authored over 30 high quality papers, and coauthored over 90 papers.



Yuanyuan Li received the Ph.D. degree in psychiatry from Sichuan University, Chengdu, China, in 2012.

She is currently an Attending Doctor with Mental Health Center, West China Hospital, Sichuan University, Chengdu. Her current research interests include psychometric analysis and data mining.

Haiqing Zhang received the Ph.D. degree in computer science and technology from the Decision and Information Sciences for Production Systems Laboratory, University Lyon 2, Lyon, France.

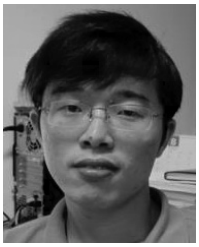
She is currently a Lecturer with the Chengdu University of Information Technology, Chengdu, China. She has published over 20 research papers in the above areas and plays an active role in the International Federation of Information Processing Ph.D. Network. She also holds one National Natural Science Foundation of China. Her current research interests include PLM maturity models, decision-making methodology, and fuzzy mathematics and data mining.

Ping Huang received the Ph.D. degree in special historical studies from Sichuan University, Chengdu, China.

She is a Professor with the School of Management, Chengdu University of Information Technology, Chengdu. She has authored over 50 papers and she participated in several projects supported by the National Natural Science Foundation of China. Her current research interest includes big data management.

Nan Han received the M.S. and Ph.D. degrees in formulas of Chinese medicine from the Chengdu University of Traditional Chinese Medicine, Chengdu, China.

She is a Lecturer with the School of Management, Chengdu University of Information Technology. She has authored over 20 papers and she participated in several projects supported by the National Natural Science Foundation of China. Her current research interests include trajectory prediction and data mining.



Rong-Hua Li received the Ph.D. degree in computer science and technology from the Chinese University of Hong Kong, Hong Kong, in 2013.

He is currently an Associate Professor with the Beijing Institute of Technology, Beijing, China. His current research interests include algorithmic aspects of social network analysis, graph data management and mining, as well as sequence data management and mining.