Selective Value Difference Metric

Chaoqun Li and Hongwei Li
Department of Mathematics, China University of Geosciences, Wuhan 430074, China
{chqli, hwli}@cug.edu.cn

***Abstract*— Value Difference Metric (VDM) is one of the widely used distance functions to define the distance between a pair of instances with nominal attributes only. Many approaches have been proposed to improve the performance of VDM. In this paper, we focus on the attribute selection approach and propose another improved Value Difference Metric. We call it Selective Value Difference Metric (SVDM). In order to learn SVDM, we investigate the attribute independence assumption held by VDM and then single out two effective attribute selection methods for SVDM. The experimental results on 36 UCI benchmark datasets validate the effectiveness of the proposed SVDM.**

***Index Terms*— Value Difference Metric; Attribute Independence Assumption; Selective Value Difference Metric; Attribute Selection; Naive Bayes.**

I. Introduction

Distance-related algorithms generally show good performance because they build different models for different query instances, and the performance of distance-related algorithms are closely related with the used distance functions. Therefore, in order to meet different data mining applications, numerous distance functions have been presented. Among which, Value Difference Metric (VDM) [1] is one of the widely used distance functions to define the distance between a pair of instances only with nominal attribute values. In order to improve the performance of VDM, researchers have done some remarkable work, such as Wilson and Martinez [2] provide nonparametric density techniques in determining the probability values of VDM. Our previous work [3] proposes One Dependence Value Difference Metric by representing the attribute dependence relationships among the attributes.

Distance metric learning is especially sensitive to the curse of dimensionality problem, and the attribute selection is one of the useful approaches to overcome it. In this paper, we focus on the attribute selection approach and propose an improved Value Difference Metric. We call it Selective Value Difference Metric (SVDM). To learn SVDM, a key problem is how to design an appropriate attribute selection method. Namely, what kind of attribute selection method is suitable for SVDM? According to our observation, VDM makes the attribute independence assumption as well as naive Bayes (NB). So our idea is that the attribute selection method for SVDM should select an attribute subset among which the attribute independence assumption is held as well as possible. In this perspective, we investigate the attribute independence assumption in VDM and then single out two effective attribute selection methods for SVDM.

Classification is one of the very important tasks in data mining. However, in many real-world applications, classification is not enough. For example, in target marketing, we may hope to evaluate various offer propositions by the class probability that a customer will respond to an offer [4]. For another example, in cost-sensitive learning, the class probability estimation is used to minimize the conditional risk [5]. Thus, accurate class probability estimation is required instead of classification in many data mining applications. For the task of class probability estimation, the conditional log likelihood (CLL) [6], [7] instead of the classification accuracy is a more reasonable performance measure for the built classifiers. Besides, in the paper by Hall [8], the mean root relative squared error (RRSE) of the probability estimates is used as a performance measure to evaluate the probability-based classifiers. Thus, in this paper, we also use CLL and RRSE to evaluate the related algorithms. The experimental results on 36 UCI benchmark datasets validate the effectiveness of our proposed SVDM.

The rest of this paper is organized as follows. In Section II, we investigate the attribute independence assumption in VDM in detail. In Section III, we propose Selective Value Difference Metric (SVDM) and single out two effective attribute selection methods for SVDM. In Section IV, we report the experimental setup and results. In Section V, we draw conclusions and outline the main directions for our future work.

II. The Attribute Independence Assumption

A. The Attribute Independence Assumption in Naive Bayes

Bayesian networks are a kind of classical probability-based models which can be used to compute the posterior probability of class c for a query instance x. In this paper, an instance x is denoted by an attribute vector $< a_1(x), a_2(x), \cdots, a_n(x) >$ and an output class c. A Bayesian network is a directed acyclic graph, and each node corresponds to a variable. The arcs between nodes represent the dependence relations between variables. The parent nodes of the ith attribute are denoted by π_i, then it estimates the posterior probabilities $P(c|x)$ by Equation 1.

$$P(c|x) = \frac{P(c)P(x|c)}{\sum P(c)P(x|c)}, \tag{1}$$

where $P(x|c) = \prod_{i=1}^{n} P(a_i(x)|c, \pi_i)$.

Therefore, accurate estimation of the conditional probabilities $P(x|c)$ is crucial to estimate the posterior probabilities. Unfortunately, fully estimation of $P(x|c)$ is an

doi:10.4304/jcp.8.9.2232-2238

NP-hard problem [9]. In order to simplify the model, naive Bayes (NB) makes an unrealistic assumption that the attributes are independent given the class. In naive Bayes, each attribute node has the class node as its parent, but has no parent from attribute nodes. This assumption is the well-known attribute independence assumption. Thanks to this assumption, naive Bayes estimates the probabilities $P(x|c)$ by Equation 2.

$$P(x|c) = \prod_{i=1}^{n} P(a_i(x)|c). \quad (2)$$

Although this assumption is rarely held in real-world applications, NB shows surprising classification performance. In order to improve the performance of NB, researchers have presented some techniques including structure expanding [10], [11], attribute selection [5], [12], [13], attribute weighting [8] and combining objective function etc [6], [7], [14].

B. The Attribute Independence Assumption in Value Difference Metric

In order to decide the difference between each pair of instances only with nominal attribute values, Stanfill and Waltz [1] present Value Difference Metric(VDM). In VDM, the difference between the ith attribute value of instance x and y is defined as

$$VDM(a_i(x), a_i(y)) = \sum_{j=1}^{k} |P(c_j|a_i(x)) - P(c_j|a_i(y))|, \quad (3)$$

where k is the number of output classes, $P(c_j|a_i(x))$ is the conditional probability that the output class of x is c_j given that the attribute A_i has the value $a_i(x)$, $P(c_j|a_i(y))$ is the conditional probability that the output class of y is c_j given that the attribute A_i has the value $a_i(y)$. In VDM, the difference between two values of an attribute is considered to be closer if they have more similar correlation with the output classes. Then VDM defines the distance between instance x and y as the sum of the value differences across all attributes. A simplified version of VDM, without the weighting schemes, can be defined as:

$$VDM(x, y) = \sum_{i=1}^{n} \sum_{j=1}^{k} |P(c_j|a_i(x)) - P(c_j|a_i(y))|. \quad (4)$$

Kasif et al. [15] explain VDM from another perspective. They point that VDM performs a simple transform on the space of the attributes, and the transformation follows from a simple probabilistic model. When the number of classes is k, a n-dimensional input space is transformed to kn-dimensional space, i.e., the ith attribute value $a_i(x)$ is transformed to a discrete probability distribution $< P(c_1|a_i(x)), \cdots, P(c_k|a_i(x)) >$, and accordingly the instance x is transformed to $< P(c_1|a_1(x)), \cdots, P(c_k|a_1(x)), \cdots, P(c_1|a_n(x)), \cdots, P(c_k|a_n(x)) >$.

They call the transformation of nominal attributes into probability distribution as probability memory based reasoning transform (probability MBR transform). According to their observation, the basis of the transformation is conditional independence assumption of the joint probability distribution, which can be defined by Equation 5.

$$P(a_1(x), \cdots, a_n(x), c) = P(c) \prod_{i=1}^{n} P(a_i(x)|c). \quad (5)$$

III. Selective Value Difference Metric

From Equation 5, the attribute independence assumption in VDM is just same as the attribute independence assumption in naive Bayes. With regards to naive Bayes, a large number of attribute selection methods are proposed to improve naive Bayes and have demonstrated remarkable performance. In addition, the attribute selection approach is also very important to distance metric learning, since the distance metric learning is especially sensitive to the curse of dimensionality problem when there are a lot of redundant and/or irrelevant attributes. In this section, we try to improve VDM by conducting an attribute selection step on the original VDM. We call it Selective Value Difference Metric (SVDM).

Selective Value Difference Metric (SVDM) is an improved VDM and it only uses the selected attribute subset from the whole attribute space to compute the distance between a pair of instances. In other words, SVDM uses Equation 6 to replace Equation 4.

$$VDM(x, y) = \sum_{i=1}^{m} \sum_{j=1}^{k} |P(c_j|a_i(x)) - P(c_j|a_i(y))|, \quad (6)$$

where m is the number of selected attributes.

To learn SVDM, a key problem is how to design an appropriate attribute selection method. Namely, what kind of attribute selection method is suitable for SVDM? Before answer this question, we first simply look back the attribute selection approach.

The attribute selection approach has attracted many researchers and many attribute selection methods have been presented. These attribute selection methods can be broadly divided into two main categories: wrapper methods [5], [12], [16], [17] and filter methods [18]–[22]. Wrapper methods depend on a learning algorithm to measure the merit of the selected attribute subsets (for example, to classification problem, the accuracy of a classifier is used to evaluate the selected attribute subsets). John and Kohavi [16] discuss the strength and weakness of wrapper methods in detail, and their experiments show that wrapper methods obtain significant improvement in classification accuracy. As for filter methods, the attribute selection process is independent of the learning algorithms, i.e., the attribute selection process can be viewed as a data preprocessing step. Generally, filter methods can be operated faster, however wrapper methods have higher accuracy than filter methods. Even so, both filters and wrappers can be encompassed within a common architecture [23]. In this architecture, attribute subset search

and attribute subset evaluation are two most important parts.

SVDM also employs this architecture to deal with the attribute selection process. Please note that, in this paper, we would like to focus more attention on the different attribute subset evaluation methods. With regards to the attribute subset search methods, we empirically employ the well-known greedy search in our forthcoming experiments. Now, the only question left to answer is what kind of attribute subset evaluation method is appropriate for SVDM.

According to our observation, VDM makes the attribute independence assumption as well as naive Bayes (NB). So our idea is that the attribute selection methods for SVDM should select an attribute subset among which the attribute independence assumption is held as well as possible. In this perspective, correlation based feature selection (CFS) [20] may be a good choice for SVDM. CFS employs some kinds of search methods, such as greedy search, to select an attribute subset and then uses Equation 7 to evaluate the merit of the selected attribute subset S containing m attributes.

$$Merit_s = \frac{m\overline{r_{ac}}}{\sqrt{m + m(m-1)\overline{r_{aa}}}}, \tag{7}$$

where $\overline{r_{ac}}$ is the average attribute-class correlation, and $\overline{r_{aa}}$ is the attribute-attribute inter-correlation. The heuristic merit tries to search an attribute subset with bigger $\overline{r_{ac}}$ by removing irrelevant attributes and smaller $\overline{r_{aa}}$ by removing redundant attributes. Smaller $\overline{r_{aa}}$ means that the attribute independence assumption is held approximately on the attribute subsets.

Besides, the naive Bayes' performance-based attribute subset evaluation method may be another good choice for SVDM. In order to improve naive Bayes, selective Bayesian classifiers (SBC) [12] is proposed. SBC uses naive Bayes' classification accuracy to evaluate alternative attribute subsets and considers adding each unselected attribute which can improve naive Bayes' classification accuracy at most on each iteration. The attribute subset evaluation method by SBC maximizes the classification accuracy of the learned naive Bayes. However, in this paper, our aim is accurate class probability estimation, so its improved algorithm called SBC-CLL [5] should be better. In SBC-CLL, the CLL of naive Bayes is used to evaluate alternative attribute subsets and considers adding each unselected attribute which can improve naive Bayes' CLL at most on each iteration. Therefore, the attribute subset evaluation method by SBC-CLL maximizes the CLL of the learned naive Bayes, and so SBC-CLL just meets our needs. SBC-CLL uses Equation 8 to evaluate the CLL of the learned selective naive Bayes on a set of test instances.

$$CLL = \sum_{i=1}^{t} logP_G(c_i|x_i), \tag{8}$$

where t is the number of test instances, c_i is the true class label of the ith test instance x_i.

In the next section, we design a group of experiments to validate our idea. For detail, we use above CFS and SBC-CLL as the attribute selection methods to select attribute subsets for SVDM. The attribute subset search method in both of them is greedy search. The experimental results on 36 UCI benchmark datasets validate the effectiveness of our proposed SVDM.

IV. Experimental Methodology and Results

In this section, we design a group of experiments to validate the effectiveness of our SVDM in terms of the conditional log likelihood (CLL) and the mean root relative squared error (RRSE) of the probability estimates produced by the distance weighted k-nearest neighbor algorithm (KNNDW). KNNDW is one of the most representative distance-related learning algorithms, in which the distance function is used twice. For detail, KNNDW first uses a distance function to find the k-nearest neighbors of the query instance, and then each of the k nearest neighbors is weighted in terms of its distance to the query instance. KNNDW should be the first test bed to demonstrate the effectiveness of a distance function.

We implement KNNDW using VDM, SVDM-CFS, and SVDM-SBC-CLL respectively in Weka platform [24] and set the number of k to be 10. Where VDM denotes Value Difference Metric, SVDM-CFS denotes Selective Value Difference Metric with CFS as the attribute selection method, and SVDM-SBC-CLL denotes Selective Value Difference Metric with SBC-CLL as the attribute selection method.

In our experiments, 36 UCI benchmark datasets published on the main web site of Weka platform [24] are used, which represent a wide range of domains and data characteristics. In our experiments, we adopted the following four preprocessing steps, which are totally same to the preprocessing steps in our previous paper [25], [26].

1) Replacing missing attribute values: We used the unsupervised filter named *ReplaceMissingValues* in Weka to replace all missing attribute values in each data set. *ReplaceMissingValues* replaces all missing values with the modes of the value of nominal attributes and means of the value of continuous attributes from the training data.
2) Discretizing numeric attribute values: We used the unsupervised filter named *Discretize* in Weka to discretize all numeric attribute values in each data set. *Discretize* discretizes all numeric attribute values using the unsupervised ten-bin discretization technique.
3) Removing useless attributes: Apparently, if the number of values of an attribute is almost equal to the number of instances in a data set, it is a useless attribute. Thus, we used the unsupervised filter named *Remove* in Weka to remove this type of attributes. In these 36 data sets, there are only three such attributes: the attribute "Hospital Number" in the data set "colic.ORIG", the attribute "instance

name" in the data set "splice" and the attribute "animal" in the data set "zoo".

4) Sampling large data sets: For saving the time of running experiments, we used the unsupervised filter named *Resample* with the size of 20% in Weka to randomly sample each large data set having no less than 5000 instances. In these 36 data sets, there are only three such data sets: "letter", "mushroom", and "waveform-5000".

Table I shows the CLL comparisons for KNNDW using VDM, SVDM-CFS, and SVDM-SBC-CLL. Experimental results on each dataset is obtained via 10 runs of 10-fold cross-validation. Runs with the various algorithms are carried out on the same training sets and evaluated on the same test sets. In particular, the cross-validation folds are the same for all the experiments on each data set. The symbols ○ and ● in the table respectively denote statistically significant upgradation or degradation over VDM with a corrected paired two-tailed t-test with the $p = 0.05$ significance level [27]. Besides, the average values and $w/t/l$ values are summarized at the bottom of the table. Each entry $w/t/l$ in the table means that, compared to VDM, SVDM-CFS and SVDM-SBC-CLL win on w datasets, tie on t datasets, and lose on l datasets.

TABLE I.
CLL OF KNNDW COMPARISONS FOR VDM VERSUS SVDM-CFS AND SVDM-SBC-CLL.

Dataset	VDM	SVDM-CFS	SVDM-SBC-CLL
anneal	-48.68±1.66	-24.79± 3.67 ○	-38.21± 2.44 ○
anneal.ORIG	-53.71±2.57	-42.37± 3.76 ○	-50.25± 2.71 ○
audiology	-54.87±2.16	-38.68± 3.80 ○	-49.16± 2.65 ○
autos	-31.98±1.45	-24.91± 2.02 ○	-25.36± 1.80 ○
balance-scale	-30.73±1.83	-30.73± 1.83	-30.73± 1.83
breast-cancer	-16.51±2.04	-16.58± 2.02	-16.90± 2.09
breast-w	-13.27±1.55	-13.02± 1.56 ○	-9.08± 2.76 ○
colic	-17.39±1.79	-17.06± 3.10	-16.32± 3.00
colic.ORIG	-18.04±1.85	-17.47± 2.43	-17.72± 2.25
credit-a	-25.66±3.29	-27.19± 4.76	-25.57± 3.99
credit-g	-52.50±2.72	-54.58± 3.83 ●	-52.20± 3.32
diabetes	-38.17±3.44	-37.90± 4.42	-37.23± 4.10
glass	-27.71±1.96	-25.49± 2.30 ○	-27.68± 2.01
heart-c	-28.08±2.16	-19.79± 2.79 ○	-23.84± 2.40 ○
heart-h	-22.42±2.51	-15.13± 2.92 ○	-20.24± 2.86 ○
heart-statlog	-11.91±1.41	-11.38± 2.49	-11.23± 1.76 ○
hepatitis	-6.28±1.14	-5.84± 1.43 ○	-6.15± 1.49
hypothyroid	-136.92±4.79	-91.03± 5.74 ○	-107.27± 5.85 ○
ionosphere	-19.39±0.67	-10.59± 2.97 ○	-10.91± 2.05 ○
iris	-4.37±1.08	-3.25± 1.24 ○	-3.28± 1.29 ○
kr-vs-kp	-50.88±3.88	-81.85± 8.45 ●	-46.91± 4.35 ○
labor	-2.16±0.42	-2.25± 1.31	-2.09± 0.54
letter	-986.33±8.97	-822.76±11.45 ○	-915.05±13.42 ○
lymph	-12.71±1.16	-9.60± 1.82 ○	-10.80± 1.44 ○
mushroom	-18.10±0.49	-13.06± 5.26 ○	-11.32± 1.84 ○
primary-tumor	-85.96±2.59	-82.89± 2.99 ○	-85.76± 2.62 ○
segment	-181.28±4.84	-105.24± 6.32 ○	-150.70± 5.56 ○
sick	-43.07±4.41	-35.49± 7.81 ○	-33.83± 7.44 ○
sonar	-13.91±0.32	-11.18± 1.94 ○	-13.02± 0.54 ○
soybean	-134.29±3.54	-109.07± 4.44 ○	-109.25±10.72 ○
splice	-289.62±2.94	-89.69± 7.88 ○	-238.20± 5.55 ○
vehicle	-84.57±2.08	-69.68± 4.59 ○	-69.57± 3.29 ○
vote	-10.19±1.61	-7.52± 3.53 ○	-6.42± 2.68 ○
vowel	-167.63±2.83	-138.10± 6.49 ○	-166.98± 2.87 ○
waveform-5000	-95.81±1.44	-76.24± 1.97 ○	-84.99± 2.10 ○
zoo	-8.23±1.03	-5.77± 0.97 ○	-7.38± 0.95 ○
Average	-78.98	-60.78	-70.32
w/t/l	-	26/8/2	26/10/0

○, ● statistically significant upgradation or degradation

Table II shows the RRSE comparisons for KNNDW using VDM, SVDM-CFS, and SVDM-SBC-CLL. Please note that, the meaning of the t-test results in this table is opposite to those in Table I. Because for the RRSE, a small number is better than a large number, which is opposite to CLL. For example, ● represents statistically significant better than VDM in terms of RRSE and ○ represents worse.

TABLE II.
RRSE OF KNNDW COMPARISONS FOR VDM VERSUS SVDM-CFS AND SVDM-SBC-CLL.

Dataset	VDM	SVDM-CFS	SVDM-SBC-CLL
anneal	72.71± 1.84	56.77± 5.64 ●	61.02± 3.05 ●
anneal.ORIG	80.50± 2.69	77.11± 5.36	77.56± 3.12 ●
audiology	98.95± 0.97	86.42± 3.89 ●	95.70± 1.37 ●
autos	96.36± 1.69	85.73± 3.15 ●	86.27± 2.86 ●
balance-scale	65.04± 2.74	65.04± 2.74	65.04± 2.74
breast-cancer	96.53± 7.14	96.32± 6.96	97.69± 7.20
breast-w	43.55± 4.52	43.01± 4.59	36.78± 8.84 ●
colic	79.25± 5.85	78.62± 9.43	76.60± 9.15
colic.ORIG	83.92± 5.88	83.72± 7.13	83.47± 6.99
credit-a	66.55± 5.96	69.06± 7.75	66.89± 6.82
credit-g	91.26± 3.07	93.21± 3.90	90.94± 3.65
diabetes	84.69± 4.85	84.58± 5.87	83.62± 5.55
glass	91.45± 2.95	88.91± 3.71 ●	91.32± 3.04
heart-c	95.48± 4.21	78.84± 7.10 ●	87.22± 5.41 ●
heart-h	88.54± 6.11	76.88± 9.07 ●	84.28± 7.46 ●
heart-statlog	73.84± 6.39	71.54±10.17	71.60± 7.76
hepatitis	86.73±10.35	83.36±12.31 ●	86.25±12.73
hypothyroid	101.59± 2.17	89.21± 2.52 ●	92.87± 2.65 ●
ionosphere	89.02± 2.10	58.74±11.59 ●	60.12± 8.59 ●
iris	40.17± 9.01	33.64±11.72 ●	33.73±11.90 ●
kr-vs-kp	35.76± 2.72	55.95± 3.65 ○	35.16± 2.95
labor	68.13± 9.71	66.90±26.60	66.70±12.18
letter	94.06± 0.24	88.39± 0.40 ●	91.96± 0.42 ●
lymph	91.28± 4.96	77.35± 9.18 ●	82.80± 6.80 ●
mushroom	21.82± 1.08	25.16± 7.79	17.51± 4.32 ●
primary-tumor	98.76± 0.87	97.80± 0.98 ●	98.70± 0.88 ●
segment	63.02± 1.03	47.13± 1.97 ●	56.07± 1.35 ●
sick	61.36± 6.13	66.24± 6.75 ○	61.50± 7.65
sonar	97.73± 0.66	84.53± 9.52 ●	93.29± 2.15 ●
soybean	90.84± 0.83	83.66± 1.39 ●	83.76± 3.28 ●
splice	93.06± 0.59	44.52± 3.09 ●	82.22± 1.26 ●
vehicle	84.04± 1.17	79.28± 2.53 ●	76.25± 2.05 ●
vote	49.20± 6.12	39.56±13.28 ●	37.66±10.47 ●
vowel	89.14± 0.59	83.05± 1.65 ●	89.00± 0.60 ●
waveform-5000	92.36± 0.85	80.47± 1.38 ●	85.95± 1.33 ●
zoo	67.80± 4.81	54.25± 5.44 ●	63.29± 4.73 ●
Average	78.46	71.53	73.63
w/t/l	-	2/12/22	0/13/23

○, ● statistically significant upgradation or degradation

From our experimental results, we can see that our SVDM-CFS and SVDM-SBC-CLL significantly outperform VDM in terms of CLL and RRSE. Now, the highlights are summarized as follows.

1) In terms of CLL of KNNDW, SVDM-CFS and SVDM-SBC-CLL both significantly outperform VDM with 26 wins and 2 losses, 26 wins and 0 losses, respectively.
2) In terms of RRSE of KNNDW, SVDM-CFS and SVDM-SBC-CLL both significantly outperform VDM with 22 wins and 2 losses, 23 wins and 0 losses, respectively.
3) SVDM-CFS and SVDM-SBC-CLL can improve the performance of VDM in domains that involve correlated attributes without reducing its performance in domains that don't. For example, it is observed that there exist many correlated attributes in the dataset "vote" [12]. As a result, SVDM-CFS and SVDM-SBC-CLL significantly outperform VDM on this dataset.

4) Generally speaking, in terms of CLL and RRSE, SVDM-CFS almost ties SVDM-SBC-CLL. However, when the computational cost is also important, SVDM-CFS should be considered firstly. From the latter experimental results, we can see that the training time complexity of SVDM-CFS is much lower than that of SVDM-SBC-CLL.

Besides, we also compare VDM with our SVDM-CFS and SVDM-SBC-CLL in terms of training time and test time. The training time and test time are the averaged CPU times in millisecond. Our experiments are performed on dual-processor 2.93 Ghz Pentium 4 Windows computer with 2Gb RAM. The detailed compared results are shown in Table III and Table IV. From the experimental results, we can see that SVDM-CFS and SVDM-SBC-CLL need some additional training time to execute the attribute selection step, compared to the original VDM. However, they can dramatically reduce the test time with 25 wins and 1 losses, 26 wins and 0 losses, respectively. Many distance-related algorithms are lazy learning algorithms, and so the test time is major indicator to evaluate their computational cost. In this perspective, our proposed SVDM-CFS and SVDM-SBC-CLL are more efficient than the original VDM.

TABLE III.
TRAINING TIME OF KNNDW COMPARISONS FOR VDM VERSUS SVDM-CFS AND SVDM-SBC-CLL.

Dataset	VDM	SVDM-CFS	SVDM-SBC-CLL
anneal	0.63±3.10	12.72±7.31 ○	4109.04± 439.19 ○
anneal.ORIG	1.74±4.98	8.92±7.79 ○	3204.51± 257.32 ○
audiology	0.79±3.46	5.15±7.38	11071.29± 569.76 ○
autos	0.15±1.50	3.59±6.61	263.13± 24.14 ○
balance-scale	0.16±1.60	1.08±3.96	6.83± 7.75 ○
breast-cancer	0.00±0.00	0.95±3.78	18.87± 6.43 ○
breast-w	0.00±0.00	3.60±6.62	29.21± 5.77 ○
colic	0.79±3.46	1.72±4.92	232.55± 15.44 ○
colic.ORIG	0.47±2.69	3.44±6.51	522.22± 22.01 ○
credit-a	0.62±3.05	1.26±4.30	129.69± 8.80 ○
credit-g	1.26±4.30	5.31±7.44	576.85± 14.31 ○
diabetes	0.61±3.00	1.90±5.17	27.30± 6.90 ○
glass	0.00±0.00	1.40±4.48	22.68± 7.90 ○
heart-c	0.48±2.74	1.88±5.12	66.91± 7.69 ○
heart-h	0.00±0.00	0.78±3.42	66.07± 7.31 ○
heart-statlog	0.00±0.00	1.08±3.96	48.49± 4.95 ○
hepatitis	0.00±0.00	2.34±5.60	86.88± 9.50 ○
hypothyroid	2.51±5.78	20.86±7.38 ○	6045.42± 474.04 ○
ionosphere	0.47±2.69	6.04±7.60	672.02± 71.77 ○
iris	0.15±1.50	0.15±1.50	1.71± 4.89
kr-vs-kp	2.65±5.89	25.63±7.52 ○	11383.94± 385.36 ○
labor	0.00±0.00	0.94±3.74	36.26± 7.33 ○
letter	1.72±4.92	27.68±6.60 ○	4048.72± 200.11 ○
lymph	0.00±0.00	0.78±3.42	97.16± 7.26 ○
mushroom	1.24±4.23	4.73±7.27	698.10± 18.95 ○
primary-tumor	0.16±1.60	3.29±6.42	383.26± 7.83 ○
segment	1.56±4.71	14.66±3.75 ○	1512.64± 26.38 ○
sick	2.98±6.19	15.48±3.52 ○	3914.53± 226.68 ○
sonar	0.32±2.25	15.01±5.45 ○	6711.73± 277.67 ○
soybean	0.78±3.42	20.96±7.38 ○	4901.26± 340.82 ○
splice	4.47±7.03	64.40±5.10 ○	65066.85±1667.57 ○
vehicle	1.24±4.23	4.34±7.00	300.61± 13.50 ○
vote	0.16±1.60	1.24±4.23	71.55± 8.12 ○
vowel	0.79±3.46	3.13±6.30	332.17± 10.56 ○
waveform-5000	0.95±3.78	30.95±1.64 ○	5753.95± 92.79 ○
zoo	0.00±0.00	0.96±3.82	65.64± 6.30 ○
Average	0.83	8.84	3680.00
w/t/l	-	11/25/0	35/1/0

○, ● statistically significant upgradation or degradation

In our another group of experiments, we use another distance-related algorithm, the 5-nearest neighbor algorithm (5NN), to compare VDM with our SVDM-CFS and SVDM-SBC-CLL in terms of the CLL, RRSE, training time, and the test time, and almost get same conclusions. The detailed experimental results can be found from Tables V - VIII.

TABLE IV.
TEST TIME OF KNNDW COMPARISONS FOR VDM VERSUS SVDM-CFS AND SVDM-SBC-CLL.

Dataset	VDM	SVDM-CFS	SVDM-SBC-CLL
anneal	603.55± 7.95	116.07± 14.38 ●	295.47± 39.57 ●
anneal.ORIG	591.13± 6.38	79.37± 14.57 ●	223.87± 19.51 ●
audiology	248.12± 8.07	22.21± 7.76 ●	81.87± 7.45 ●
autos	25.30± 7.58	8.59± 7.82 ●	9.36± 7.69 ●
balance-scale	21.54± 7.60	21.10± 7.53	21.73± 7.59
breast-cancer	7.42± 7.77	4.18± 6.91	6.28± 7.74
breast-w	41.57± 7.46	40.34± 7.75	17.37± 4.97 ●
colic	27.64± 6.62	4.68± 7.19 ●	10.26± 7.41 ●
colic.ORIG	31.88± 3.13	5.00± 7.33 ●	20.94± 7.42 ●
credit-a	63.43± 3.73	25.25± 7.55 ●	32.15± 6.22 ●
credit-g	175.31± 6.52	33.42± 8.59 ●	136.08± 10.27 ●
diabetes	44.84± 5.33	18.29± 6.30 ●	24.36± 7.85 ●
glass	10.80± 7.29	6.97± 7.75	10.17± 7.51
heart-c	21.71± 7.69	10.77± 7.27 ●	17.62± 6.16
heart-h	21.06± 7.54	4.82± 7.23 ●	14.70± 3.76 ●
heart-statlog	9.22± 7.73	4.96± 7.27	6.08± 7.65
hepatitis	4.55± 7.16	2.35± 5.63	2.05± 5.33
hypothyroid	5800.47±10.58	4299.28±221.73 ●	2999.27±279.29 ●
ionosphere	38.76± 7.85	5.30± 7.43 ●	10.50± 7.42 ●
iris	1.26± 4.30	0.79± 3.46	0.47± 2.69
kr-vs-kp	3139.22± 8.36	1087.49± 29.47 ●	2140.95±151.33 ●
labor	0.63± 3.10	0.00± 0.00	0.00± 0.00
letter	21606.91±22.57	12313.59± 18.97 ●	16702.69±606.25 ●
lymph	6.11± 7.69	3.44± 6.51	4.22± 6.98
mushroom	509.21± 7.73	161.25± 8.29 ●	174.64± 8.12 ●
primary-tumor	134.42± 7.74	92.29± 9.68 ●	126.22± 4.16 ●
segment	2353.48± 7.70	679.41± 7.84 ●	1514.06± 38.18 ●
sick	3514.68± 7.66	3891.40± 88.84 ○	1684.84± 83.92 ●
sonar	23.93± 7.86	3.12± 6.28 ●	13.98± 4.71 ●
soybean	945.86± 9.09	598.21± 38.83 ●	607.34± 54.75 ●
splice	7408.22± 5.56	782.64± 5.03 ●	4327.74±183.65 ●
vehicle	195.23± 7.91	46.47± 12.49 ●	90.15± 6.59 ●
vote	27.83± 6.50	7.83± 7.88 ●	8.65± 7.87 ●
vowel	484.08± 3.13	80.47± 5.64 ●	408.45± 25.81 ●
waveform-5000	489.52± 7.40	171.45± 3.56 ●	335.32± 12.01 ●
zoo	4.22± 6.98	2.15± 5.36	4.35± 7.02
Average	1350.92	684.30	891.23
w/t/l	-	1/10/25	0/10/26

○, ● statistically significant upgradation or degradation

V. CONCLUSIONS AND FUTURE WORK

In this paper, we propose an improved Value Difference Metric by conducting an attribute selection step on the original VDM. We call it Selective Value Difference Metric (SVDM). To learn SVDM, we investigate the attribute independence assumption in VDM and then find that CFS and SBC-CLL are two effective attribute selection methods for SVDM.

From the experimental results, we can see that our proposed SVDM-CFS and SVDM-SBC-CLL significantly outperform VDM in terms of the conditional log likelihood (CLL) and the mean root relative squared error (RRSE). Although SVDM-CFS and SVDM-SBC-CLL need some additional training time to execute the attribute selection step, they dramatically reduce the test time of the original VDM.

In our current version, for simplicity, we limit the attribute subset search method to greedy search. In fact, however, there exist of many other search methods, such as genetic search, BestFirst search, random search, and so

TABLE V.
CLL OF 5NN COMPARISONS FOR VDM VERSUS SVDM-CFS AND SVDM-SBC-CLL.

Dataset	VDM	SVDM-CFS	SVDM-SBC-CLL
anneal	-54.94±1.14	-25.14±3.58 ○	-48.92±2.39 ○
anneal.ORIG	-62.91±2.67	-44.00±4.01 ○	-59.71±2.77 ○
audiology	-45.37±2.14	-40.30±4.36 ○	-44.97±2.31
autos	-24.06±2.19	-22.55±2.12 ○	-22.70±2.44 ○
balance-scale	-34.78±1.79	-34.78±1.79	-34.78±1.79
breast-cancer	-17.29±1.86	-16.71±2.05	-17.21±1.91
breast-w	-12.55±1.87	-12.61±1.94	-10.26±2.59 ○
colic	-16.65±2.68	-17.06±3.10	-17.08±2.88
colic.ORIG	-17.99±2.40	-17.09±2.45	-17.93±2.50
credit-a	-27.49±3.09	-27.19±4.76	-26.45±3.63
credit-g	-54.45±3.67	-54.85±3.86	-54.32±3.41
diabetes	-39.78±3.65	-38.73±4.09	-38.95±4.05
glass	-26.66±2.17	-26.33±2.29	-27.03±2.26
heart-c	-23.66±2.30	-21.75±2.81 ○	-24.20±2.39
heart-h	-23.47±2.34	-16.21±2.80 ○	-23.08±2.42
heart-statlog	-11.56±1.72	-11.78±2.33	-11.55±1.87
hepatitis	-6.46±1.44	-6.31±1.55	-6.64±1.64
hypothyroid	-169.80±4.76	-91.31±5.73 ○	-117.15±6.39 ○
ionosphere	-10.63±2.23	-10.86±2.71	-10.63±2.18
iris	-5.30±1.05	-3.99±1.14 ○	-4.02±1.22 ○
kr-vs-kp	-64.61±3.39	-81.85±8.45 ●	-57.59±5.30 ○
labor	-1.76±0.85	-2.29±1.35	-1.83±0.87
letter	-773.79±8.75	-754.47±9.59 ○	-778.37±9.12 ●
lymph	-10.20±1.61	-9.98±1.79	-10.16±1.74
mushroom	-25.25±0.96	-13.14±5.25 ○	-14.55±1.83 ○
primary-tumor	-84.01±2.60	-83.56±3.25	-83.93±2.63
segment	-173.10±4.26	-116.56±6.46 ○	-167.54±4.63 ○
sick	-55.62±3.96	-35.56±7.81 ○	-34.38±7.33 ○
sonar	-9.35±1.86	-11.53±2.42 ●	-9.35±1.85
soybean	-100.88±2.30	-98.10±2.51 ○	-98.11±3.46 ○
splice	-123.91±5.19	-109.03±7.33 ○	-122.05±5.29
vehicle	-69.29±3.15	-70.39±4.36	-72.94±3.47 ●
vote	-9.27±1.76	-7.52±3.53 ○	-6.49±2.65 ○
vowel	-119.94±2.88	-140.12±6.59 ●	-120.65±3.02 ●
waveform-5000	-56.51±3.44	-55.31±3.77	-61.16±4.76 ●
zoo	-7.11±0.74	-5.43±1.07 ○	-6.12±0.85 ○
Average	-65.84	-59.29	-62.86
w/t/l	-	16/17/3	13/19/4

○, ● statistically significant upgradation or degradation

TABLE VI.
RRSE OF 5NN COMPARISONS FOR VDM VERSUS SVDM-CFS AND SVDM-SBC-CLL.

Dataset	VDM	SVDM-CFS	SVDM-SBC-CLL
anneal	79.31± 1.21	57.10± 5.23 ●	73.21± 2.57 ●
anneal.ORIG	88.06± 2.54	78.55± 4.46 ●	85.53± 2.75 ●
audiology	93.64± 1.00	88.14± 4.53 ●	93.30± 1.15
autos	84.52± 3.57	81.61± 3.85 ●	81.68± 4.50 ●
balance-scale	69.56± 2.69	69.56± 2.69	69.56± 2.69
breast-cancer	99.43± 6.51	96.87± 7.45	98.98± 6.79
breast-w	40.00± 5.96	40.15± 6.09	38.77± 7.89
colic	77.50± 8.28	78.64± 9.40	78.89± 8.78
colic.ORIG	83.96± 7.24	81.98± 7.38	83.69± 7.64
credit-a	69.29± 5.44	69.06± 7.75	68.04± 6.15
credit-g	93.17± 3.95	93.60± 3.84	92.97± 3.67
diabetes	86.73± 5.07	85.84± 5.52	85.89± 5.50
glass	89.43± 3.43	89.45± 3.54	90.01± 3.53
heart-c	86.77± 5.17	82.97± 6.65 ●	87.87± 5.36
heart-h	91.10± 5.51	78.78± 8.06 ●	90.41± 5.82
heart-statlog	72.70± 7.61	73.59± 9.41	72.78± 8.01
hepatitis	87.97±13.08	86.92±13.81	89.25±14.34
hypothyroid	115.32± 1.93	89.32± 2.52 ●	97.11± 2.63 ●
ionosphere	58.68± 9.70	59.86±10.29	58.62± 9.31
iris	45.80± 7.81	38.55± 9.85 ●	38.67±10.11 ●
kr-vs-kp	40.05± 2.27	55.95± 3.65 ○	38.19± 2.87 ●
labor	56.71±21.64	66.98±28.13	58.35±21.67
letter	87.87± 0.26	86.41± 0.34 ●	87.92± 0.27
lymph	80.07± 7.78	78.67± 8.89	79.72± 8.55
mushroom	29.27± 1.58	25.22± 7.74	20.64± 3.38 ●
primary-tumor	97.94± 0.79	97.77± 1.06	97.90± 0.81
segment	61.33± 0.92	50.13± 1.85 ●	60.08± 1.03 ●
sick	69.42± 5.14	66.28± 6.75 ●	61.38± 7.57 ●
sonar	74.98± 9.92	85.89±11.18 ○	75.01± 9.95
soybean	82.48± 0.62	81.28± 0.79 ●	81.35± 1.21 ●
splice	52.36± 1.82	49.48± 2.65 ●	51.68± 1.87
vehicle	75.44± 2.03	79.40± 2.43 ○	77.76± 2.15 ○
vote	44.54± 7.21	39.56±13.28	38.26± 9.96 ●
vowel	76.92± 0.88	83.49± 1.66 ○	77.13± 0.92 ○
waveform-5000	67.72± 2.92	66.65± 3.23	71.38± 3.70 ○
zoo	62.42± 3.66	52.95± 6.93 ●	56.61± 4.72 ●
Average	74.24	71.85	72.46
w/t/l	-	4/18/14	3/21/12

○, ● statistically significant upgradation or degradation

on. Therefore, thoroughgoingly researching these search methods is a main direction for our future work.

ACKNOWLEDGEMENTS

We thank the anonymous reviewers for their very useful comments. The work was supported by the National Natural Science Foundation of China under grant no. 61203287 and 61071188, the Provincial Natural Science Foundation of Hubei under grant no. 2012FFB6401 and 2011CDA103, and the Fundamental Research Funds for the Central Universities under grant no. CUG130414.

REFERENCES

[1] C. Stanfill and D. Waltz, "Toward memory-based reasoning," *Communications of the ACM*, vol. 29, pp. 1213–1228, 1986.

[2] D. R. Wilson and T. R. Martinez, "Improved heterogeneous distance functions," *Journal of Artificial Intelligence Research*, vol. 6, pp. 1–34, 1997.

[3] C. Li and H. Li, "One dependence value difference metric," *Knowledge-Based Systems*, vol. 24, pp. 589–594, 2011.

[4] M. Saar-Tsechansky and F. Provost, "Active sampling for class probability estimation and ranking," *Machine Learning*, vol. 54, pp. 153–178, 2004.

[5] L. Jiang and H. Zhang, "Learning naive bayes for probability estimation by feature selection," in *Proceedings of the 19th Canadian Conference on Artificial Intelligence*. Springer Press, 2006, pp. 503–514.

[6] D. Grossman and P. Domingos, "Learning bayesian network classifiers by maximizing conditional likelihood," in *Proceedings of the Twenty-First International Conference on Machine Learning*. ACM Press, 2004, pp. 361–368.

[7] D. Lowd and P. Domingos, "Naive bayes models for probability estimation," in *Proceedings of the Twenty-Second International Conference on Machine Learning*. ACM Press, 2005, pp. 529–536.

[8] M. Hall, "A decision tree-based attribute weighting filter for naive bayes," *Knowledge-Based Systems*, vol. 20, pp. 120–126, 2007.

[9] D. M. Chickering, "Learning bayesian networks is np-complete," in *Learning from Data: Artificial Intelligence and Statistics*. Springer Press, 1996, pp. 121–130.

[10] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol. 29, pp. 131–163, 1997.

[11] L. Jiang, Z. Cai, D. Wang, and H. Zhang, "Improving tree augmented naive bayes for class probability estimation," *Knowledge-Based Systems*, vol. 26, pp. 239–245, 2012.

[12] P. Langley and S. Sage, "Induction of selective bayesian classifiers," ser. Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence, 1994, pp. 339–406.

[13] C. A. Ratanamahatana and D. Gunopulos, "Feature selection for the naive bayesian classifier using decision trees," *Applied Artificial Intelligence*, vol. 17, pp. 475–487, 2003.

[14] L. Jiang and C. Li, "Scaling up the accuracy of decision-tree classifiers: A naive-bayes combination," *Journal of Computers*, vol. 6, pp. 1325–1331, 2011.

[15] S. Kasif, S. Salzberg, D. Waltz, J. Rachlin, and D. Aha,

TABLE VII.
TRAINING TIME OF 5NN COMPARISONS FOR VDM VERSUS SVDM-CFS AND SVDM-SBC-CLL.

Dataset	VDM	SVDM-CFS	SVDM-SBC-CLL
anneal	0.79±3.46	13.26±6.81 ○	4129.39± 444.25 ○
anneal.ORIG	0.47±2.69	9.09±7.78 ○	3226.74± 258.21 ○
audiology	0.15±1.50	8.74±7.79 ○	11144.73± 582.85 ○
autos	0.00±0.00	4.97±7.29	268.28± 23.70 ○
balance-scale	0.00±0.00	0.95±3.78	6.24± 7.69 ○
breast-cancer	0.15±1.50	0.46±2.63	18.24± 5.93 ○
breast-w	0.64±3.15	2.82±6.05	27.04± 7.67 ○
colic	0.00±0.00	1.40±4.48	234.53± 17.46 ○
colic.ORIG	0.31±2.18	3.60±6.62	537.01± 28.05 ○
credit-a	0.31±2.18	2.05±5.33	133.14± 12.62 ○
credit-g	0.79±3.46	5.19±7.44	583.44± 15.39 ○
diabetes	0.15±1.50	1.89±5.15	27.51± 6.73 ○
glass	0.15±1.50	1.24±4.23	22.79± 7.85 ○
heart-c	0.00±0.00	2.82±6.05	67.79± 8.07 ○
heart-h	0.00±0.00	1.87±5.09	66.14± 6.63 ○
heart-statlog	0.16±1.60	2.35±5.63	48.89± 5.32 ○
hepatitis	0.16±1.60	2.75±5.90	88.79± 9.32 ○
hypothyroid	3.92±6.83	19.32±6.73 ○	6131.53± 482.48 ○
ionosphere	0.16±1.60	5.60±7.51 ○	677.98± 71.92 ○
iris	0.00±0.00	0.47±2.69	1.72± 4.92
kr-vs-kp	3.93±6.84	24.54±7.77 ○	11433.33± 385.73 ○
labor	0.00±0.00	0.78±3.42	35.64± 7.11 ○
letter	2.37±5.67	26.78±7.16 ○	4038.27± 86.51 ○
lymph	0.15±1.50	0.63±3.10	96.68± 6.86 ○
mushroom	1.08±3.96	7.22±7.87	703.73± 20.68 ○
primary-tumor	0.16±1.60	4.16±6.88	387.48± 6.29 ○
segment	0.94±3.74	14.91±4.11 ○	1522.19± 26.80 ○
sick	3.59±6.61	15.98±2.28 ○	3930.74± 226.66 ○
sonar	0.32±2.25	14.39±5.33 ○	6716.97± 279.02 ○
soybean	0.47±2.69	21.10±7.48 ○	4923.78± 344.72 ○
splice	5.47±7.50	65.54±6.28 ○	65227.33±1680.93 ○
vehicle	0.16±1.60	5.30±7.43	301.62± 13.06 ○
vote	0.15±1.50	0.48±2.74	70.79± 9.35 ○
vowel	0.46±2.63	3.92±6.83	331.13± 10.40 ○
waveform-5000	1.08±3.96	31.97±3.12 ○	5774.50± 94.38 ○
zoo	0.00±0.00	1.72±4.92	66.36± 6.84 ○
Average	0.80	9.17	3694.51
w/t/l	-	13/23/0	35/1/0

○, ● statistically significant upgradation or degradation

TABLE VIII.
TEST TIME OF 5NN COMPARISONS FOR VDM VERSUS SVDM-CFS AND SVDM-SBC-CLL.

Dataset	VDM	SVDM-CFS	SVDM-SBC-CLL
anneal	591.36± 6.85	102.39± 12.85 ●	291.69± 39.99 ●
anneal.ORIG	580.57± 6.22	74.37± 17.56 ●	219.93± 20.89 ●
audiology	235.98± 9.27	19.69± 7.58 ●	79.71± 6.90 ●
autos	24.38± 7.78	6.90± 7.83 ●	8.28± 7.84 ●
balance-scale	19.37± 6.73	20.45± 7.27	20.47± 7.24
breast-cancer	6.70± 7.76	3.15± 6.33	6.27± 7.72
breast-w	38.89± 7.88	37.96± 7.82	15.75± 4.14 ●
colic	26.39± 7.30	4.53± 7.13 ●	12.32± 6.40 ●
colic.ORIG	30.95± 2.26	4.52± 7.11 ●	20.47± 7.30 ●
credit-a	63.11± 3.15	21.71± 7.69 ●	32.50± 4.32 ●
credit-g	175.31± 6.57	31.36± 9.30 ●	136.62± 8.52 ●
diabetes	44.38± 5.77	18.13± 5.86 ●	22.96± 7.84 ●
glass	10.32± 7.45	4.06± 6.89	9.21± 7.73
heart-c	20.79± 7.39	8.26± 7.83 ●	15.34± 3.84
heart-h	19.67± 6.90	4.99± 7.32 ●	14.53± 4.03
heart-statlog	8.59± 7.82	3.28± 6.40	7.33± 7.83
hepatitis	4.20± 6.95	2.38± 5.70	1.55± 4.68
hypothyroid	5715.27±12.28	4109.88±217.29 ●	2946.09±286.25 ●
ionosphere	35.78± 7.15	4.37± 7.05 ●	9.81± 7.57 ●
iris	1.26± 4.30	0.79± 3.46	0.62± 3.05
kr-vs-kp	3093.61± 9.76	1025.40± 27.77 ●	2139.33±152.78 ●
labor	0.63± 3.10	0.16± 1.60	0.32± 2.25
letter	21631.81±38.16	12323.03± 19.25 ●	16718.74±602.85 ●
lymph	6.58± 7.78	2.52± 5.81	4.52± 7.11
mushroom	504.18± 6.91	151.41± 8.28 ●	170.34± 7.89 ●
primary-tumor	129.03± 6.84	88.41± 8.34 ●	121.90± 6.24 ●
segment	2353.55±10.84	669.55± 5.56 ●	1499.82± 35.27 ●
sick	3441.15± 8.05	3812.21± 87.41 ○	1543.56± 71.81 ●
sonar	22.49± 7.80	3.27± 6.38 ●	13.30± 5.63 ●
soybean	931.54± 9.13	591.22± 38.18 ●	597.33± 52.31 ●
splice	7375.77± 5.66	781.25± 3.88 ●	4254.68±182.43 ●
vehicle	192.67± 7.39	42.98± 12.28 ●	87.67± 7.66 ●
vote	26.09± 7.39	7.33± 7.83 ●	7.95± 7.84 ●
vowel	480.79± 6.59	78.29± 1.53 ●	405.07± 24.99 ●
waveform-5000	480.94± 6.56	170.42± 4.37 ●	326.17± 11.59 ●
zoo	3.46± 6.55	1.26± 4.30	3.29± 6.42
Average	1342.43	673.11	882.37
w/t/l	-	1/10/25	0/11/25

○, ● statistically significant upgradation or degradation

"A probabilistic framework for memory-based reasoning," *Artificial Intelligence*, vol. 104, pp. 287–311, 1998.

[16] G. John and R. Kohavi, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, pp. 273–324, 1997.

[17] L. Jiang, Z. Cai, H. Zhang, and D. Wang, "Not so greedy: Randomly selected naive bayes," *Expert Systems with Applications*, vol. 39, pp. 11 022–11 028, 2012.

[18] R. John, G. Kohavi, and K. Perfleg, "Using decision trees to improve cased-based learning," in *Proceedings of the tenth international conferencce on machine learning*. Morgan Kaufmann, 1993, pp. 25–32.

[19] ——, "Irrelevant features and the subset selection problem," in *Proceedings of the eleventh international conferencce on machine learning*. Morgan Kaufmann, 1994, pp. 121–129.

[20] M. Hall, "Correlation-based feature selection for discrete and numeric class machine learning," in *Proceedings of the 17th International Conference on Machine Learning*. Morgan Kaufmann, 2000, pp. 359–366.

[21] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE trasaction on pattern analysis and machine intelegence*, vol. 27, pp. 1226–1238, 2005.

[22] Z. Wu, J. Zhang, and Y. Gao, "An attribute reduction algorithm based on genetic algorithm and discernibility matrix," *Journal of Software*, vol. 7, pp. 2640–2648, 2012.

[23] P. N. Tan, M. Steinbach, and V. Kumar, *Introduction to data mining*, 1st ed. Pearson Education, 2006.

[24] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*, 2nd ed. San Francisco: Morgan Kaufmann, 2005.

[25] C. Li and H. Li, "A modified short and fukunaga metric based on the attribute independence assumption," *Pattern Recognition Letters*, vol. 39, pp. 1213–1218, 2012.

[26] ——, "A survey of distance metrics for nominal attributes," *Journal of Software*, vol. 7, pp. 1262–1269, 2010.

[27] C. Nadeau and Y. Bengio, "Inference for the generalization error," *Machine Learning*, vol. 52, pp. 239–281, 2003.

Chaoqun Li is currently a lecturer in Department of Mathematics at China University of Geosciences (Wuhan). Her research interests include data mining and machine learning.

Hongwei Li is currently a professor in Department of Mathematics at China University of Geosciences (Wuhan). His research interests include machine learning and information process.