Cost-sensitive Naïve Bayes Classification of Uncertain Data

Xing Zhang[1]; Mei Li[1]; Yang Zhang[1*]; Jifeng Ning [2]
1. College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling 712100, China
2. College of Information Engineering, Northwest A&F University, Yangling 712100, China
Email: zhangyang@nwsuaf.edu.cn

***Abstract*—Data uncertainty is widespread in real-word applications. It has captured a lot of attention, but little job has been paid to the research of cost sensitive algorithm on uncertain data. The paper proposes a novel cost-sensitive Naïve Bayes algorithm CS-DTU for classifying and predicting uncertain datasets. In the paper, we apply probability and statistics theory on uncertain data model, define the utility of uncertain attribute to total cost, and propose a new test strategy for attribute selection algorithm. Experimental results on UCI Datasets demonstrate the proposed algorithm can effectively reduce total cost, and significantly outperforms the competitor.**

***Index Terms*—Uncertain data, Cost sensitive, Naïve Bayes**

I. Introduction

Most existing data mining algorithms requires as input certain and precision data. In many real-life applications, however, data contains inherent uncertainty due to many reasons, such as the random nature of the physical data generation and collection process, measurement and decision errors, unreliable data transmission and data staling. The technology of mining uncertain data has attracted great interest, and many works have been done in the field.

Classification models for uncertain data are evaluated by their classification precision so far. The aim of these classification models is to minimize classification errors. When classification errors are proportional to the cost of misclassifications, minimizing classification errors can lead to the minimum cost of classification. however in the real world, this might be wrong. Thus researchers propose the cost-sensitive learning with the objective to minimize the cost. Inductive learning methods that consider a variety of costs are often referred to as cost-sensitive learning, but few can handle uncertain data.

In the case of medical diagnosis, to diagnose the disease of a patient, a doctor must decide whether a medical test is worthwhile to perform and if so, which one. Each test can improve the accuracy of diagnosis, and meanwhile bring a certain amount of test cost. Misdiagnoses also bring misclassification cost. Our goal is minimize the sum of test cost and misclassification cost. And in this process, the data may be uncertain.

In this paper, we proposed a cost-sensitive naïve Bayes classifier for uncertain data(CS-UNB). We extend the test-cost sensitive naive Bayes classification (CSNB)[1] to uncertain data. On the basis of the frame of CSNB, we defined the influence of uncertain attribute to total cost, propose a new measurement for attribute selection based on the influence, and use the method in NBU[2] to handle data uncertainty. In experiment, we compare CS-UNB with CSDTU[3] on UCI datasets. Experimental result shows that CS-UNB has a better performance and more robust than the competitor.

This paper is organized as follows. Section 2 reviews the related work. Section 3 describes the problem definition of our work. Section 4 presents the training and testing algorithm of CS-UNB. Section 5 gives the experiment result and discussion. And the concludes of this paper and our future work is given in section 6.

II. Related Work

A. Data Uncertainty

The concept of the data uncertainty was investigated quite intensively in recent years, many works have been done in classification, the common practice for uncertain data classification is expending the traditional algorithm. Bi and Zhang proposed a total support vector classification algorithm (TSVC)[4]. The formulation of support vector classification with uncertain input data was motivated by the total least squares regression method. In 2007, Yang proposed a USVC algorithm[5] which extent the support vector classification by incorporating input uncertainties. In the same year, he presented a iterative approach AUSVC[6] ,which combined TSVC and USVC to achieve a better performance. In 2009, Qin proposed a decision tree based classification method on uncertain data (DTU)[7], which considered the uncertain data interval and probability distribution function(PDF). In [8], Tsang raised another uncertain decision tree based classification UDT, and came up with a few strategies for pruning candidate split points. In [9], Ren presented a novel naïve Bayes classification algorithm for uncertain data with a PDF. The key solution was to extend the class conditional probability estimation in the Bayes model to handle PDF. Qin proposed another uncertain naïve Bayes based

doi:10.4304/jcp.9.8.1897-1903

classification algorithm NBU[2], which applied probability and statistics theory on uncertain data model, and provide solutions for model parameter estimation for both uncertain numerical data and uncertain categorical data.

B. Cost-sensitive Learning

The aim of above algorithms is to minimize the misclassification error, unlike cost-sensitive algorithms' goal is to minimize the total cost. There are a mounts of works on cost-sensitive researching. In [10], Turney analyzed a whole variety of costs in machine learning, , and two types of costs were considered as the most important: the misclassification costs and the test costs. Some previous works, for example [11], only considers misclassification costs, while [12] only considers the test costs, and they all partial. The best way is to minimize both the misclassification costs and the test costs. There are many works have been done for this. In [13], Turney proposed a system called ICET, which used a genetic algorithm to build a decision tree aiming at minimize the cost of tests and misclassifications. In [14], Ling proposed a new decision tree learning algorithm using minimum total cost of tests and misclassifications as the attribute split criterion.

A few works have been done on Cost-sensitive Naïve Bayes Classification. In [15], Lizotte studied the theoretical aspects of active learning with test costs based on naïve Bayes classifiers. In [1], Chai presented a cost-sensitive learning algorithm called CSNB. The test strategy of CSNB determines how unknown attributes are selected to perform test on in order to minimize the sum of the misclassification costs and test costs.

C. Cost-sensitive Learning on Uncertain Data

At the time of this writing, few extensions have been made to consider Cost-sensitive learning with data uncertainty. In [3], Liu proposed a method extending traditional cost-sensitive decision tree to uncertain data. Because a decision tree places different levels of importance on the attributes by the natural organization of the tree, it cannot be easily fitted to make flexible decisions on selecting unknown attributes for tests. However, the naïve Bayes based algorithms overcome these difficulties more naturally. In this paper, we focus on providing a Naïve Bayes based Classification to handle Cost-sensitive learning on Uncertain Data.

III. Problem Definition

In this section, we give a formal definition of the problem, for simplicity, we only consider binary classification and only handle categorical attribute, but our work is easy to extend it to numerical attribute and multi-classification.

We write $A=\{A_1,A_2,\cdots,A_M\}$ for the set of attributes, and $C=\{c_1,c_2,\cdots,c_{|c|}\}$ for the set of class labels. Here, each attribute $A_i \in A$ could be either a certain attribute or an uncertain attribute. We write A_i for a certain attribute. The value of A_i , denoted by v_k, is a value from a domain $Dom(A_i)=\{v_1,v_2,\cdots,v_n\}$. And we write A_i^u for an uncertain attribute, whose value is characterized by probability distribution over domain $Dom(A_i^u)$. It can be represented by the probability vector $P=\{P_{i1},P_{i2},\cdots,P_{in}\}$, we write A_{ij}^u for a certain attribute, the j-*th* value of the probability vector.

In cost-sensitive learning, classifying a new sample on a new case, we often consider the test cost, denoted by $cost_{test}$, when missing values must be obtained through physical test which incur costs themselves. Suppose that T_j is a sample of dataset D, each attribute A_i can be either known or unknown. Let $\tilde{A}$ denote the set of known attributes among all the attributes A and $\bar{A}$ the unknown attributes, $A=\tilde{A}\cup\bar{A}$, $cost_{test}(\tilde{A})=0$, and $cost_{test}(\bar{A})>0$.

Once a classifier is built, it gives a sample T_j to be classified a class label c_j. While the correct class label of T_j is c_i, the misclassification incurs costs, and we call that misclassification cost. Suppose that $cost_{ij}$ is the cost of predicting a sample of class c_i as belonging to class c_j, it is clear that $cost_{ii}=0$, and $cost_{ij}\neq cost_{ji}$.

The total cost is the sum of the test cost $cost_{test}$ and the misclassification cost $cost_{ij}$. We construct a classifier with the aim to minimize the total cost for uncertain data.

IV. The Proposed Scheme

We introduce the CS-UNB algorithm in two procedures: the test strategy for each test case with the aim to minimize the total cost and each step of the algorithm in detail.

A. Test Strategy

In this section, we mainly illustrate the test strategy of CS-UNB classifier. When a new test sample with missing values comes, a CS-UNB classifier needs a test strategy to decide unknown attributes should be selected for testing. The test strategy is aimed to minimize the sum of the misclassification cost , and test cost and it can handle uncertain data.

During the process of classification, based on the results of previous tests, decisions are made sequentially on whether a further test on an unknown attribute should be performed, and if so, which attribute to select. More specifically, the selection of the next unknown attribute to test is not only dependent on all the values of initially known attributes, but also dependent on the values of those unknown attributes previously tested. [16]

A test brings a certain amount of test cost $cost_{test}$. Meanwhile it may reduce the misclassification cost $cost_{ij}$. If the reduction of misclassification is larger than the increase of test cost, the test is helpful. In order to decide whether a test is helpful and which attribute should be selected, we write $Util(\bar{A}_i)$ represent the utility of attribute $\bar{A}_i$ to total cost, $\bar{A}_i\in\bar{A}$.

$$Util(\bar{A}_i) = Gain(\tilde{A}, \bar{A}_i) - c_{test}(\bar{A}_i) \qquad (1)$$

Here, $c_{test}(\bar{A}_i)$ represents the test cost of $\bar{A}_i$. $Gain(\tilde{A}, \bar{A}_i)$ is the reduction in the expected misclassification cost obtained from knowing $\bar{A}_i$'s true value, which is given by:

$$Gain(\tilde{A}, \bar{A}_i) = c_{mc}(\tilde{A}) - c_{mc}(\tilde{A} \cup \{\bar{A}_i\}) \quad (2)$$

$c_{mc}(\tilde{A})$ is the misclassification cost from $\tilde{A}$, $c_{mc}(\tilde{A})=\min R(c_j \mid \tilde{A})$. $c_{mc}(\tilde{A} \cup \{\bar{A}_i\})$ is the expected misclassification cost from $\tilde{A} \cup \{\bar{A}_i\}$. Since the value of $\bar{A}_i$ is not revealed until the test is performed, we calculate it by considering expectation over all possible values of $\bar{A}_i$ as follows:

$$c_{mc}(\tilde{A} \cup \{\bar{A}_i\}) = \sum_{k=1}^{|Dom(\bar{A}_i)|} P\left(\bar{A}_i = v_k \middle| \tilde{A}\right)$$

$$\times \min_{c_j \in c} R(c_j \mid \tilde{A}, \bar{A}_i = v_k) \quad (3)$$

Here, $P(\bar{A}_i=v_k|\tilde{A})$ is the conditional probability of $\bar{A}_i=v_k$ premised on $\tilde{A}$, and $R(c_j|\tilde{A}, \bar{A}_i=v_k)$ is misclassification cost with the class label c_j premised on $\tilde{A}$ and $\bar{A}_i=v_k$.

$R(c_j|\tilde{A})$ is easily obtained using Equation(4):

$$R(c_j \mid \tilde{A}) = \sum_{k=1}^{|C|} \cos t_{kj} \times P\left(c_k \middle| \tilde{A}\right), 1 \le k \le |C| \quad (4)$$

$cost_{kj}$ is the cost of predicting a sample of c_k class as belonging to class c_j. $P(c_k|\tilde{A})$ represents the conditional probability with the class label c_k premised on $\tilde{A}$. According to the Bayesian theory, we have:

$$P\left(c_k \middle| \tilde{A}\right) = \frac{P\left(\tilde{A} \middle| c_k\right) P(c_k)}{P\left(\tilde{A}\right)} \quad (5)$$

Here, $P(c_k)$ and $P(\tilde{A})$ are constants. According to Bayesian assumption, we have:

$$P(\tilde{A} \mid c_k) = \prod_{A_i^u \in \tilde{A}} P(A_i^u \mid c_k) \quad (6)$$

To calculate A_i^u, that is:

$$P(A_i^u \mid c_k) = p_{i1} P(A_{i1}^u \mid c_k) + p_{i2} P(A_{i2}^u \mid c_k) +$$

$$\cdots + p_{ij} P(A_{ij}^u \mid c_k) \quad (7)$$

It is easy to calculate $P(A_{ij}^u | c_k)$ by:

$$P(A_{ij}^u = v_m \mid c_k) = \frac{PC(v_m, c_k)}{PC(c_k)} \quad (8)$$

Here, $PC(c_k)$ represents the probabilistic cardinality of samples with class c_k in the dataset. It can be estimated below:

$$PC(c_k) = \sum_{j=1}^{|D|} P(c_{T_j} = c_k) \quad (9)$$

$PC(v_m, c_k)$ denotes the sum of the probability of each sample in class c_k whose value equals to v_m. That is,

$$PC(v_m, c_k) = \sum_{j=1}^{|D|} P(v_m \in T_j \wedge c_{T_j} = c_k) \quad (10)$$

Overall, when an attribute offers more gain than the cost it brings, it is worth testing. It means that if $Util(\bar{A}_i)>0$, a test is needed. It is easy to calculate all the $Util(\bar{A}_i)$ of testing unknown attributes in $\bar{A}$, and we select $\bar{A}_i^*$ to test($i^*=argmax_i Util(\bar{A}_i)$).

We obtain the attribute value of $\bar{A}_i^*$ after the attribute is tested. The set of known attributes is updated by $\tilde{A}=\tilde{A} \cup \{\bar{A}_i^*\}$ and correspondingly, $\bar{A}$ is updated by $\bar{A}=\bar{A}-\{\bar{A}_i^*\}$. Repeat the selection process until $Util(\bar{A}_i)$ is non-positive or there is no unknown attribute left. The expanded known attribute set $\tilde{A}$ is used to predict the class label.

Finally, the misclassification cost is $cost_{ij}$ if sample T_i predicted as class c_j is actually from class c_i. All the costs brought by the attribute tests comprise the test cost $cost_{test}$. The total cost $cost_{total} = cost_{ij} + cost_{test}$.

B. CS-UNB Algorithm

In this section, we will describe each step of the CS-UNB algorithm for constructing a Cost-sensitive Naïve Bayes Classification of Uncertain data. In the training phrase, a CS-UNB classifier is learned from the training dataset D; in the testing phase, for each test sample, a test strategy is designed to minimize the total cost based on the CS-UNB obtained.

Algorithm 1 gives us the training algorithm of algorithm of CS-UNB. Learning a CS-UNB classifier is similar to the process of estimating the distribution parameters in traditional NB.

Algorithm 1 CS-UNB Learning algorithm for CS-UNB

```
Input:  Training dataset D;
Output : Cost-sensitive uncertain naive Bayes Classier B;
Begin:
for (Each sample T_j∈D ) do
  for (each attribute A_i ) do
    if ( A_i is uncertain categorical) then
      for (each v_m∈A_i )
        PC(v_m,c_k)=updateuncertain(P(v_m∈T_j∧c_T_j=c_k))
      end for
    else if ( A_i is categorical) then
      for (each v_m∈A_i )
        PC(v_m,c_k)=updatecertain(T_j.v_m)
      end for
    end if
  end for
  PC(c_k)=updateProbalilisticCardinality(T_j.class)
end for
```

For each sample with uncertain, we update $PC(v_m, c_k)$ by adding the probability of each sample in class c_k whose value equals to v_m using function *updateuncertain()*. For each categorical sample, we update $PC(v_m, c_k)$ by the value of T_j, it can be 0 or 1. For each instance T_j, we update the probabilistic cardinality for class c_k of the dataset by the class of the sample T_j using function *updataProbalilisticCardinality()*.

The details of the classification algorithm CS-UNB are given in Algorithm 2.

Algorithm 2 Testing algorithm for CS-UNB

Input: Text example T_j, Cost-sensitive Uncertain naive Bayes Classier B;

Output : The predicted class label;

Begin:

1: Let $\tilde{A}$ = the set of known attributes. Let $\bar{A}$ = the set of unknown attributes.

2: set $\cos t_{test} = 0$ =0

3: while ($\bar{A} \notin \varnothing$) do

4: for(each $\bar{A}_i \in \bar{A}$)

5: calculate $Util(\bar{A}_i)$

6: end for

7: if (all $Util(\bar{A}_i) \leq 0$) then

8: break;

9: else $\bar{A}_i^* = \max Util(\bar{A}_i)$

10: end if

11: $\cos t_{test} = \cos t_{test} + \cos t_{test}(\bar{A}_i^*)$

12: Reveal the value of $\bar{A}_i^*$

13: $\tilde{A} = \tilde{A} \cup \{\bar{A}_i^*\}$

14: $\tilde{A} = \tilde{A} - \{\bar{A}_i^*\}$

15: end while

16: label = $\arg\min_j \left(R(c_j|\tilde{A}) + \cos t_{test}\right)$

When $\bar{A} \notin \varnothing$, for each attribute $\bar{A}_i$, $(\bar{A}_i \in \bar{A})$, we calculate $Util(\bar{A}_i)$ (step 4,5,6). For each attribute $\bar{A}_i$, $Util(\bar{A}_i) \leq 0$, there is no more tests needed(step 7,8). When a test is needed, we select the attribute with the highest $Util(\bar{A}_i)$ (step 9). Add the test cost spend on $\bar{A}_{i^*}$ to the total test cost (step 11). Move $\bar{A}_{i^*}$ from $\bar{A}$ to $\tilde{A}$ (step 13,14). Give the predicted class label(step 17).

The time complexity of CS-UNB is $O(Npq)$, where N denotes the number of samples, p denotes the number of attributes, $q = \max_{i=1}^{p} \left|Dom(X_i^u)\right|$. All values of attributes had to be searched in order to calculate the probabilistic cardinality for the value of the specific attribute. The memory required by CS-UNB are not related to the total number of samples, but dominated by the sufficient statistics. The space complexity of CS-UNB is $O(pqr)$. Here, r denotes the number of class label. In this paper, we only consider binary classification tasks, so the number of classes is 2 and the space complexity of the implementation is $O(2pq) = O(pq)$.

V. EXPERIMENTS

To validate the performance of CS-UNB algorithm, we conduct experiments on UCI dataset. The algorithms were implemented based on WEKA. All experiments are executed on a PC with Intel Cor2 Duo 2.52GHz CPU and 2.0GB main memory. A collection, containing 9 real-world benchmark datasets, with categorical attributes and binary classification tasks were assembled from the UCI Repository. The detail information is listed in TABLE 1.

TABLE I.
DATASETS USED IN EXPERIMENTS

Dataset	Attribute	Sample	Class Distribution (pos/neg)
Breast-w	10	699	458/241
Vote	17	435	267/168
Car	7	1733	1211/522
Bank	11	600	274/326
Breast-cancer	10	286	201/85
Ecoli	8	336	220/116
Heart-statlog	14	270	150/120
Kr-vs-kp	37	3196	1669/1527
Tic-tac-toe	10	985	322/626

A. Data Preprocessing

Because there is no real-life uncertain dataset publicly available yet, we need to convert existent certain data into uncertain data in our experiment, Such method is widely used by the research community[2, 7, 17]. For each attribute A_i^u, we first convert it into a probability vector $P=\{P_{i1},P_{i2},\cdots,P_{in}\}$, where P_{im} is the probability that A_{ij}^u has value v_m, that is, $P(A_{ij}^u = v_m)=P_{im}$. If the original value of A_{ij}^u is equal to v_m, we set P_{im} to be a value less than 1, and evenly distribute the rest probability 1- P_{im} to all other values, that is :

$$\sum_{k=1 \wedge k \neq m}^{n} p_{ik} = 1 - p_{im} \qquad (11)$$

For example, when we introduce 10% uncertainty, there is 90% probability that the attribute will take the original value and 10% probability to take any of other values. Suppose in the original certain dataset $A_{ij}^u = v_1$, then we will assign P_{i1}=90%, and assign $p_{ij} (2 \leq j \leq k)$ to ensure $\sum_{j=2}^{k} p_{ij} = 10\%$. We denote this dataset with 10% uncertainty by U0.1.

Test cost of attribute is assigned by random values between 0 and 100. We use FP to denote the number of positive sample which is misclassified as negative, and FN to denote the number of negative sample is misclassified as positive. The proportion of misclassification FP/FN is set to 600/1000, 1000/1000 and 1000/2000.

B. Experiment with Uncertain Level

Figure 1 compares the average total cost of CS-UNB, NBU[2] and CSDTU[3] on datasets bank, breast-w, kr-vs-kp and vote. In these experiments, the uncertain level U is set from 0 to 0.5, increasing by 0.1 each time, and FP/FN = 1000/1000.

Because of limited space, we will not list all the experimental results. Instead, we select for datasets: Bank, Breast-w, Kr-vs-kp and Vote to show how uncertain level affects total cost.

From Figure 1, it appears that with the increase of uncertainty level, NBU algorithm keeps stable and even increase, while the total cost of both CS-UNB and CSDTU drop dramatically after U0.4. As NBU is aimed to minimize misclassification error, so the tests don't change a lot with the increase. However, the amount of tests in CS-UNB and CSDTU drops when uncertainty goes up to certain extent. Few tests lead to the decrease of test cost and it has little impact to misclassification cost. So the total cost falls. We also can see, the total cost of CS-UNB is smaller than that of other except ks-vs-kp. This is further proof that CS-UNB has a better performance.

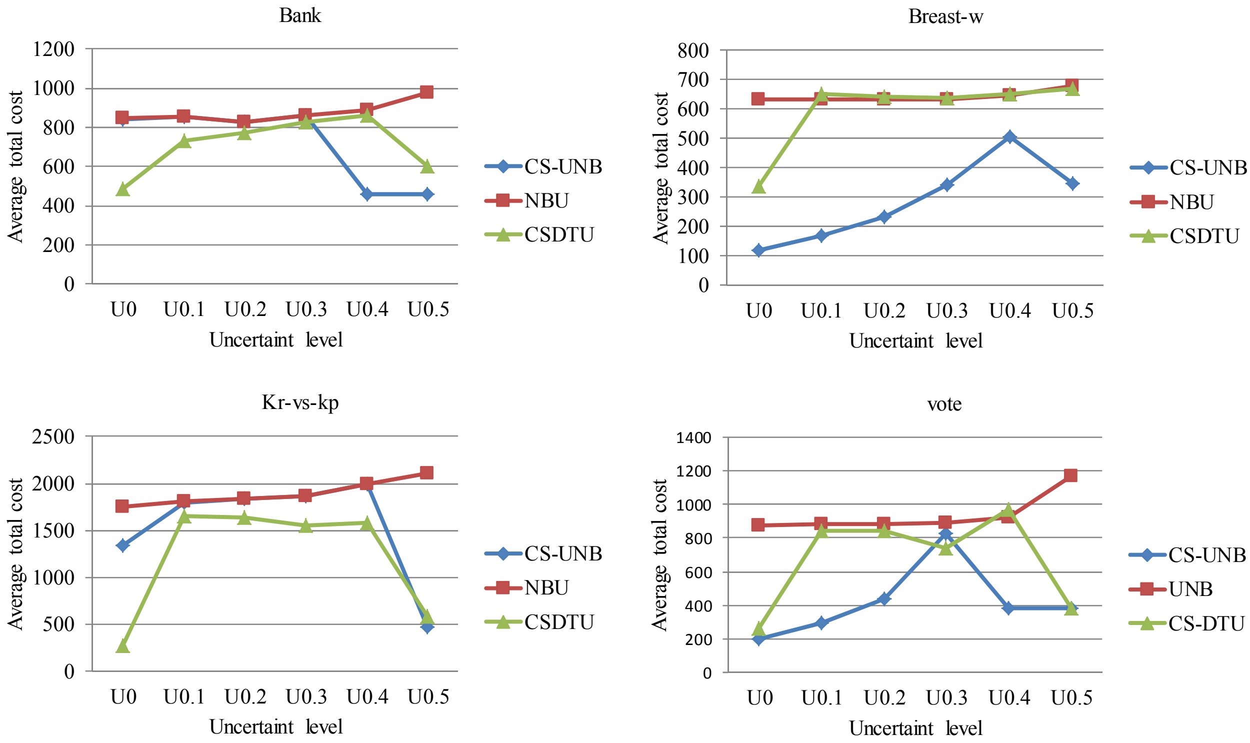

Fig.1 Comparisons with varying uncertaint level

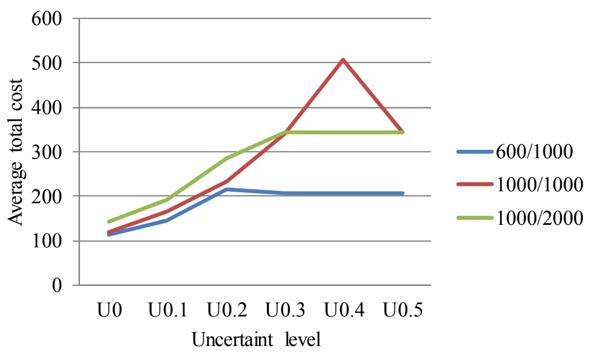

Fig.2 Comparisons with different FP/FN on dataset Breast-w

C. Experiment with FP/FN

FP is the cost of one false positive example, and FN is the cost of one false negative example. In order to study the effect of different FP/FN to CS-UNB, we select a representative dataset Breast-w to illustrate the impact of parameter FP/FN towards the algorithm performance.

Figure 2 shows the result of CS-UNB on dataset Breast-w when FP/FN is set to 600/1000, 1000/1000 and 1000/2000.

From Figure 2, the total cost is different with varied FP/FN, and has a similar trend. This is because different FP/FN brings different misclassification cost and lead to different total cost. The similar trend shows there is no great performance impact with different FP/FN.

D. Performance Comparison

Figure 3 compares the average total cost of CS-UNB, NBU[2] and CSDTU[3] on all the nine datasets. In these experiments, the percentage of uncertain is U0.20, and the proportion of misclassification FP/FN is 1000/1000.

It can be observed from Figure 3, although both CS-UNB and NBU are based on Naive Bayes model, the average total cost of CS-UNB, which is aim to minimize total cost is smaller than NBU, which is aim to minimize misclassification error. when we compare CS-UNB with CSDTU which has the same goal of minimum total cost, the results show that CS-UNB has evident advantage on dataset car, breast-cancer, breast-w, tric-tac-toe and vote. The performance of the both is similar on dataset ecoli and heart-statlog. CS-UNB performs slightly better than than CSDTU on dataset bank and kr-vs-kp. Overall, we can say, CS-UNB has better performance.

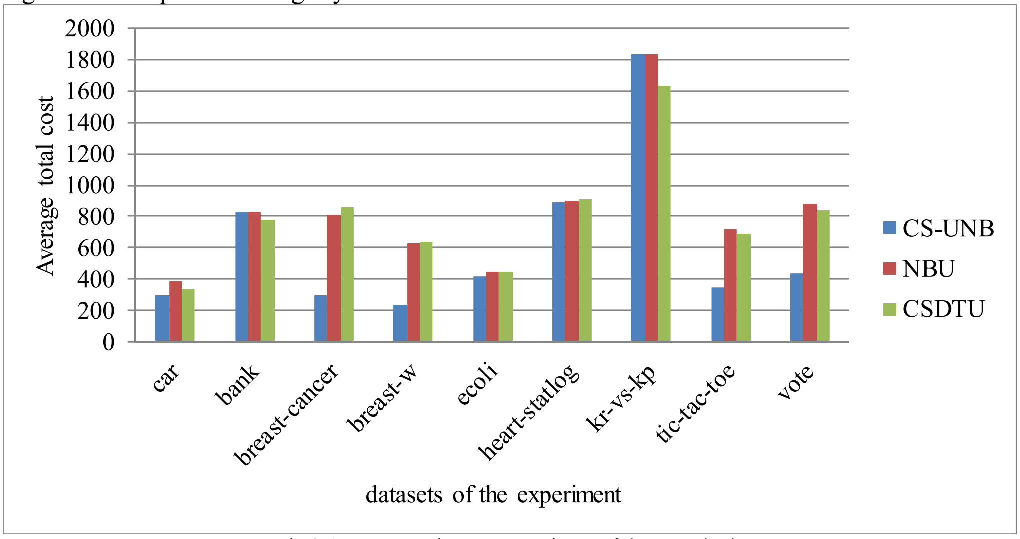

Fig.3 Average total cost comparisons of three methods

Analysing the datasets, we can see that, when a dataset has a large number of attributes, the performance of CS-UNB is decreased. This is due to the assumption on the condition independence. Specifically, the attributes are always relevant each other, so lots of attributes maybe induce performance degradation.

VI. The Proposed Scheme

In this paper, we propose a new cost-sensitive naïve bayes algorithm, namely CS-DTU, to classify for classifying and predicting uncertain datasets. We integrate the uncertain data model into cost sensitive naïve bayes algorithm. On the basis of the frame of CSNB, we define the utility of uncertain attribute to total cost, propose a new test strategy for selection of attribute. The new method allows us to derive cost sensitive model based on uncertain data and attain lower total cost. Our experimental result demonstrates that CS-UNB outperforms other competing algorithms.

In the future, we will extend the methods to uncertain numerical attributes, and generalize the ideas for performing batch tests that involve a number of tests to be done together, rather than a sequential manner.

Acknowledgment

The authors wish to thank Xiaoyan Cai, Libin Yang and Mingjian Liu. The work was supported by the Natural Science Foundation of China under Grant No. 61303125, No. 61303226 and No. 61003151.

References

[1] Chai, X., et al. *Test-cost sensitive naive bayes classification*. in *Data Mining, 2004. ICDM'04. Fourth IEEE International Conference on*. 2004. IEEE.

[2] Qin, B., Y. Xia, and F. Li. *A Bayesian classifier for uncertain data*. in *Proceedings of the 2010 ACM Symposium on Applied Computing*. 2010. ACM.

[3] Liu, M., et al., *Cost-sensitive decision tree for uncertain data*, in *Advanced Data Mining and Applications*. 2011, Springer. p. 243-255.

[4] Zhang, J.B.T., *Support vector classification with input data uncertainty.* Advances in neural information processing systems, 2005. 17: p. 161-169.

[5] Yang, J. and S. Gunn. *Exploiting uncertain data in support vector classification*. in *Knowledge-Based Intelligent Information and Engineering Systems*. 2007. Springer.

[6] Yang, J. and S. Gunn, *Iterative Constraints in Support Vector Classification with Uncertain Information.* CONSTRAINT-BASED MINING AND LEARNING, 2007: p. 49.

[7] Qin, B., Y. Xia, and F. Li, *DTU: a decision tree for uncertain data*, in *Advances in Knowledge Discovery and Data Mining*. 2009, Springer. p. 4-15.

[8] Tsang, S., et al., *Decision trees for uncertain data.* Knowledge and Data Engineering, IEEE Transactions on, 2011. 23(1): p. 64-78.

[9] Ren, J., et al. *Naive bayes classification of uncertain data*. in *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on*. 2009. IEEE.

[10] Turney, P., *Types of cost in inductive concept learning.* In proceedings of the 17th Cost-Sensitive Learning Workshop(ICML), 2000.

[11] Elkan, C. *The foundations of cost-sensitive learning*. in *International joint conference on artificial intelligence*. 2001. Citeseer.

[12] Tan, M., *Cost-sensitive learning of classification knowledge and its applications in robotics.* Machine Learning, 1993. 13(1): p. 7-33.

[13] Turney, P., *Cost-sensitive classification: Empirical evaluation of a hybrid genetic decision tree induction algorithm.* Journal of Artificial Intelligence Research (JAIR), 1995. 2.

[14] Ling, C.X., et al. *Decision trees with minimal costs*. in *Proceedings of the twenty-first international conference on Machine learning*. 2004. ACM.

[15] Ling, C.X., et al. *Decision trees with minimal costs*. in *Proceedings of the twenty-first international conference on Machine learning*. 2004. ACM.

[16] Chai X, Deng L, Yang Q, et al. *Test-cost sensitive naive bayes classification[C]//Data Mining, 2004. ICDM'04.*

Fourth IEEE International Conference on. IEEE, 2004: 51-58.
[17] Qin, B., et al. *A rule-based classification algorithm for uncertain data*. in *Data Engineering, 2009. ICDE'09. IEEE 25th International Conference on*. 2009. IEEE.

Xing Zhang, born in Jingzhou City, Hubei Province, China, on Sep. 18, 1986. Received Bachelor of Engineering, Master of Engineering from Northwest A&F University, China in 2009 and 2012 respectively. From Sep. 2012, Ph D. candidate in Agricultural Electrification and Automation from Northwest A&F University, China.

He is a PhD candidate currently. He has joined and accomplished 3 national and provincial research programs, such as "citation-based summarization of scientific papers". He has also published 2 papers in international conferences/journals. His major research interests include uncertain data mining and machine learning.

Mei Li, born in Yangling City, Shaanxi Province, P. R. China, on Dec. 9, 1981. Received Bachelor of Science in computer science and technology from Northwest University, China in 2004 and Master of Engineer in computer application technique from Northwest A&F University, China in 2009. From Sep. 2009, Ph D. candidate in Agricultural Electrification and Automation from Northwest A&F University, China.

She has been working in Northwest A&F University since 2004. She has joined national and provincial research programs, such as "citation-based summarization of scientific papers". Her major research interests include text classification and uncertain data mining.

Yang Zhang, born in Yangzhou City, Jiangsu Province, China, on Apr. 19, 1975. Received Ph D of Computer Software and Theory from Northwestern Polytechnical University, China in 2005.

He is a professor and Doctoral Supervisor at Northwest Northwest A&F University. His research interests involve data mining and machine learning.

Jifeng Ning, born in Hancheng City, Shaanxi Province, P. R. China, on Jun. 1975. Received Ph D of signal and information processing from Xidian University, China in 2009.

He is an associate professor at Northwest Northwest A&F University. His research interests involve computer vision and pattern recognition.