ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 4, Issue 5 , May 2017

Testing of Cost effective learning algorithms

Sonia Jain, Avinash Dhole

P.G. Student, Department of Computer Science and Engineering, Raipur Institute Of Technology, Raipur, Chhattisgarh, India

Sr.Assistant Professor, Department of Computer Science and Engineering, Raipur Institute Of Technology, Raipur, Chhattisgarh, India

ABSTRACT: The research focuses on decision trees that take account of the cost of acquiring attributes for decision making in many real-world applications. How to build an inexpensive and reliable inductive learning model, the decision-making process must learn which sequence to perform to accomplish its task. Many previous works have successfully reduced the total test cost in the area of test-cost sensitive decision tree learning, but also the classification accuracy simultaneously degraded. This paper works on a new idea, i.e., at the cost of the loss of classification accuracy, it does not has to reduce the total test cost. For that, a multi-target adaptive attribute selection measure will be proposed and also for building and testing decision trees, a simple but effective method will be taken. Our algorithm uses a random attribute selection measure to find an appropriate attribute to test at each node in the tree, instead of using a greedy attribute selection measure like many other decision tree learning algorithms. Specifically, through the whole space of attributes in tree building, we conduct a random search. By this way, the total test cost is reduced by the algorithm significantly and compared to its competitors, it maintains the higher classification accuracy at the same time. The effectiveness of our proposed randomly selected decision tree algorithm is validated by the experimental results on 6 UCI datasets.

I. INTRODUCTION

As a kind of inductive learning algorithm, decision tree algorithms have been successful to build classifiers with the aim to maximize the classification accuracy. The well-known ID3 [1], C4.5 [2], CART [3], and so on all center around inducing decision trees for the high classification accuracy. However, one of the main difficulties of tree building in practice is that the majority of variables tests have associated cost, which may be diverse for each test [5,6]. Since data is not free, instead of only focusing on classification accuracy, a learner should perform an economic yet effective induction in practical application. That is to say, when building decision trees on a training data or performing a test on a new instance, if the tests incur the cost themselves, we should consider the total test cost and decide if it is worthwhile to pay the test cost. Test-cost sensitive learning is more practical than simple traditional classification in many applications such as intelligent medical diagnostic systems [7]. As an example, in medical diagnosis, an expert needs to evaluate the tradeoff between the accuracy (the proportion of patients diagnosed correctly) and efficiency (the cost of measuring attribute values). Before diagnosing a patient, some tests for this patient, such as diastolic blood pressure test or serum insulin test, may not yet be known and generally take different cost. Like in the Pima Indians Diabetes dataset [8], a serum insulin test takes $22.78 for a patient while a diastolic blood pressure test only takes $1. These tests provide different informational values towards maximizing the classification accuracy, while performing them will incur extra cost. So, we have to pursue the balance between classifiers' reliability and low-cost testing.

To the best of our knowledge, some existing test-cost sensitive learning algorithms are about balancing the act of two types of cost, namely the misclassification cost and the test cost, to deter-mine which test will be done [8–13]. The others focus on the balance between classification accuracy and minimal test cost directly [14–19]. Dealing with the high-cost test classification problems, decision trees are a kind of feasible candidate. When a test case is classified by a decision tree, some algorithms [20–24] have tried to find a tradeoff between the accuracy and the test cost. These algorithms are all the improved test-cost sensitive versions based on ID3 or C4.5 and they directly adapt existing information theoretic measures by including costs. Through experiment and study, results show that, compared with C4.5, all these algorithms reduce the test cost, unfortunately, yet at the same time degrade the classification accuracy.

In this paper we focus on building decision trees which have not only the lower test cost but also the higher classification accuracy. Previous works [20–24] reduce the test cost while also degrade the classification accuracy. In the medical diagnosis and other fields, the higher classification accuracy is also one of the most important factors. This

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 4, Issue 5 , May 2017

fact raises the question of whether we can build decision trees which reach the same classification accuracy as C4.5,mean while reduce the test cost significantly. To this end, instead of using the greedy attribute selection measures employed by previous works [20–24], the randomness is introduced to the tree building to select appropriate attributes. More specifically, we carry a random search through the whole useful candidate attributes. When selecting the current attribute to build a tree, we cannot only consider the total test cost, but also the classification accuracy.

The rest of this paper is organized as follows. Section 2 introduces some related works on attribute selection measures indecision tree learning and test-cost sensitive decision tree. Section3 proposes our test-cost sensitive decision tree learning algorithm. Section 4 conducts a series of experiments on a large suite of bench-mark datasets to validate our algorithm. Section 5 concludes the paper and outlines the main directions for future study.

II. RELATED WORK

A. Attribute selection measures in decision tree learning

A decision tree consists of a tree structural model and a setof decision nodes and leaves. The structural model is a directly decision-making process in which a leave specifies a class value and decision node specifies a test over one of the attributes, called theattribute selected at the node. Attribute selection is quantified for the root node using a statistical measure given a set of examples.The examples are then filtered into subsets according to values of the selected attribute. The same process is applied recursively to each of the subsets until all nodes are leaves. Decision tree learning algorithms such as C4.5 [2] are often used for classification problems. On the base of the information gain ratio, a selection measure is utilized in C4.5, which can be defined as follows.

$$GainRatio(P,R) = \frac{Gain(P,R)}{SplitInfomation(P,R)} \qquad (1)$$

where Gain(P, R), called information gain, denotes the reduction of impurity from the parent node (before splitting) to the child nodes(after splitting). Gain(P, R) is defined as

$$Gain(P,R) = Entropy(P) - \sum_{m=1}^{j} \frac{|P_m|}{|P|} Entropy(P_m) \qquad (2)$$

where Entropy(P), called entropy, describes the purity of the given instances set, k is the number of the split attribute values, Siis thesubset of instances at the ith child node of the parent node.Splitinformation (P,R) is the split information of the selected splitattribute, which is defined as follows.

$$SplitInfomation(P,R) = -\sum_{m=1}^{j} \frac{|P_m|}{|P|} \log_2 \frac{|P_m|}{|P|} \qquad (3)$$

In addition to the information gain ratio measure, some other attribute selection measures in decision tree learning can be found from Jiang et al. [26] and Jiang [27].

Note that, such greedy attribute selection measures may have the potential to suffering from local optimum. Aiming at this problem, Breiman [4] provides a framework of random split selection for tree ensembles, which is well known as the “random forests”.It is a classifier consisting of many decision trees. Its output classis the mode of the classes output by individual trees. The randomizing variable is the key factor in the algorithm, and it is typically used in the selection of the node and coordinates to split when atree is built.

B. Test-cost sensitive decision tress.

Traditional decision tree learning algorithms such as C4.5 aim to maximize the classification accuracy. However, in many real-world applications, the cost of acquiring attribute values is diverse and expensive [14,28,15–19], and thus it is more reasonable to induce decision trees that take account of test cost of attributes. As shown in the previous subsection, top-down greedy algorithms for inducing decision trees use information theoretic measures, such as the

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 4, Issue 5 , May 2017

information gain ratio measure, to select an appropriate attribute during the tree induction process. Naturally, many scholars adapt those measures by introduce the test cost of attributes. Extended algorithms that considering the test cost include EG2 [20], IDX [21], CS-ID3 [22], CSGR [23], CS-C4.5 [24] and so on. By introducing the test cost of attributes, these works mainly focus on minimizing the total test cost and adapting information theoretic measures towards attributes that cost less. An advantage of the above adaptive decision tree learning algorithms is that it naturally extends the information theoretic measures by introducing the test cost. In Ling and Charles [9], the test-cost sensitive learning is converted as the theory of Decision Trees with Minimal Cost (DTMC). Instead of adapting the information gain to introduce the test cost, Ling and Charles [9] use the misclassification cost and the test cost directly as the cost reduction splitting criteria. Besides, Sheng et al. [29] present an approach where a decision tree is built for each new test case. For a given new case, depending on the expected cost calculated so far, the optimal policy suggests a best attribute to minimize the total costs. Their research adopts an optimal strategy, which may also have the potential to local optimum. Another related work [16] is the filter attribute selection method that takes into account the test cost of features, which proposes a framework for test-cost sensitive feature selection (CS-CFS) based on CFS (Correlation-based Feature Selection). CS-CFS consists of adding a new term to the evaluation function of a filter feature selection method so that the test cost is taken into account. It is defined as

$$MC_s = \frac{k\overline{r_{ci}}}{\sqrt{k+(k-1)\overline{r_{ii}}}} - \lambda \frac{\sum_i^k C_{test}(A_i)}{k} \quad (4)$$

where MCsis the merit of the selected attribute subset S affected by the cost of the features , k is the size of attribute subset, ̄rciis theaverage feature -class correlation, ̄riiis the average feature-featureinter-correlation, Ctest(Ai) is the test cost of the feature Ai, and _ isa parameter introduced to weight the influence of the cost in the evaluation function. Through experimentation with these algorithms, we have found that the classification accuracy of these decision tree algorithms may have not been recognized enough. To make up the disadvantage, it is the goal of this paper that building decision trees which keep high classification accuracy meanwhile reduce the total test cost significantly. Since those algorithms that adapt existing information theoretic measures by introducing the test cost [20–24]degrade the classifiers' accuracy, our algorithm is not going to adapt the existing information theoretic measures to introduce the test cost. At the same time, we also do not use the optimal strategies like Ling and Charles [9] and Sheng et al. [29] in order to avoid the potential to local optimum. In contrast, our work adopts the random selected strategy. An random factor is introduced to regulate the influence of our strategy and makes the built decision trees more biased in favor of the test cost or the classification accuracy.

III. PROPOSED ALGORITHM

In this section, we propose a test-cost sensitive decision tree learning algorithm called randomly selected decision tree which aims to pursue both the higher classification accuracy and the lower test cost. For this purpose, a random strategy, instead of a greedy strategy, is used to find an appropriate attribute for each splitting. Specifically, given an attribute set AS containing mattributes and a random factor β, an adaptive attribute selection operator is performed. Within the rang of (0, β), the best attributeAtbest, which has the highest information gain ratio among mattributes, is selected. Because Attbestis the attribute with the highest information gain ratio, in this case, the process of tree-building pays more attention to the built decision tree's accuracy. Otherwise, selects an attribute, denoted as AttProper. The process of looking for AttProperis given later. AttProperis an attribute with the lowest test cost among all candidates, in this case, the process of tree-building pays more attention to the built decision tree's test cost while it still pursues a certain accuracy. Let Attsbe the selected split attribute at the current split node, which is defined

$$Att_s = \begin{cases} Att_{best}, & if\ rand(0,1) < \beta \\ Att_{best}, & otherwise \end{cases} \quad (5)$$

Now, the only left thing is how to find AttProper. To an attribute set AS containing m attributes, we firstly perform a ranking operator, in terms of the attributes' information gain ratio, to rank all attributes in descending order. Based on the ordered attribute array AS, we can obtain an attribute subset AS= {A1,A2, . . .,A η }, which has only the top η attributes with the highest information gain ratio. Here η is defined as

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 4, Issue 5 , May 2017

$$\eta = \min\{1 + \log_2 m, g\} \quad (6)$$

where g is the number of the attributes whose information gain ratio is greater than 0.
Algorithm 1.Training (TD, AS, TC, ˇ).
Input: TD-a training dataset; AS-an attribute set; TC-an array listing the test cost of each attribute; ˇ-a random factor
Output: DT-the built test-cost sensitive decision tree1: if the number of training instances is under 2 then2: Create a leaf node for the tree3: else4: m : = size of(AS)5: for i = 1 to m do6: Calculate the ith attribute's Gain Ratio using Eq. (1)7: end for8: Sort AS in descending order of GainRatio9: if the maximum Gain Ratio is zero then 10: Create a leaf node for the tree 11: else12: if rand (0,1) < ˇ then 13: Use the attribute Attbestto split the tree 14: else15: Calculate _ using Eq. (6)16: Obtain the attribute subset ¯AS 17: Find AttProperusing ¯AS and Eq. (7)18: Use the attribute AttProperto split the tree19: end if20: Create a child node for each possible value of the split attribute21: For each child node, recursively call the algorithm22: end if23: end if24: Return a test-cost sensitive decision tree

Compared to the time complexity O(nm2) of the standard decision-tree learning algorithm C4.5 [2,30], Algorithm 1 needs some additional time to sort m attributes. The additional time complexity for sorting m attributes is only O(mlog2m), where n is the number of training instances and m is the number of attributes. Therefore, we can conclude that Algorithm 1 almost maintains computational simplicity and efficiency that characterize standard decision-tree learning algorithm C4.5.After the tree is built, the following discussion is how to deal with test instances in order to predict the class of the test instances with the minimal total test cost. In this paper, we consider the strategy to follow the tree built in the previous section. Yang [31] note that the decision trees had already specified an order in which to perform the tests. Aimed at reducing the total test cost without any loss of accuracy in the process of building a decision tree, it is reasonable to follow the test sequential. Algorithm 2 outlines thetesting algorithm of .

Algorithm 2.Testing (DT, TC, x).Input: DT-the built test-cost sensitive decision tree by Algorithm 1; TC-an array listing the test cost of each attribute; x-a test instance Output: c-the predicted class; Ttest-the total test cost for x1: Sort x down the built tree DT from the root node to someone leaf node L2: Estimate the class membership probabilities of x using the training instances dropping into the leaf node Land then predict its class label c3: According to TC, calculate the total test cost Ttest of all split attributes in this path4: Return c and Ttest

Since algorithm is inherently unstable, we stabilize the estimated class membership probabilities by building an ensemble of algorithm using bagging and averaging the estimated class membership probabilities across the ensemble like Breiman [4], Jiang [32] and Jianget al. [33]. The calculation of the average total test cost for ensemble trees is as follows: firstly, the average total test cost for a single tree is calculated by Ttestof all testing instances. We will then calculate the average total test cost for ensemble decision trees. Bagging has two parameters: the number of bagging iterations and the percent-age of the training data to use for learning in each iteration. In our experiments, we use the parameter settings with 30 and100, respectively. To our knowledge, Hall [34] stabilizes the estimated attribute weights by building multiple decision trees usingbagging. Ahmad [35] creates ensembles of decision trees such that.

Table I Classification Accuracy% comparison .

	Accuracy			
Dataset	Classifier			
	c4.5	Random forest	Naïve bayes	AdaBoost
vowel	83.131	97.47	67.07	17.37
vehicle	72.58	76.01	44.79	40.187
glass	69.199	80.33	48.55	44.87
letter	88.04	96.53	64.11	70.9
iris	94.66	94.66	94	96
Breast cancer	72.63	71.13	72.94	73.45

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 4, Issue 5 , May 2017

IV. EXPERIMENTAL ANALYSIS

The main purpose of carrying these operations is to validate the effectiveness of the algorithm in terms of accuracy, precision, recall and auc. We have used all 6 UCI datasets published on the main website of weka platform. In our experimentation we used the same preprocessing steps on these datasets and then we replaced the missing values with mode and means of available data. Numeric attributes values are discretized using the unsupervised ten-bin discretization implemented on WEKA. We can see the values of the comparison of accuracy, precision, recall and AUC in the Table I, Table II, Table III and table IV respectively.

Table II Classification of precision values on the following algorithms.

	Precision			
Dataset	Classifier			
	c4.5	Random forest	Naïve bayes	AdaBoost
Vowel	94.46	1	85.4	0
Vehicle	86.62	87.89	39.54	43.03
Glass	73.94	78.5	47.9	45.66
Letter	90.39	99.45	77.61	0
Iris	1	1	1	1
breast cancer	75.08	75.82	79.56	77.13

Table III Classification of recall values on the following algorithms.

	Recall			
Dataset	Classifier			
	C4.5	Random forest	Naïve bayes	AdaBoost
Vowel	92.22	97.77	77.77	0
Vehicle	89.42	95.97	87.44	1
Glass	80	90	82.85	1
Letter	90.32	97.95	60.63	0
Iris	98	1	1	1
breast cancer	91.84	87.71	83.68	89.31

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 4, Issue 5 , May 2017

Table IV Classification of AUC% on the following algorithms.

	AUC			
Dataset	Classifier			
	c4.5	Random Forest	Naïve bayes	AdaBoost
Vowel	96.52	1	96.9	56.66
Vehicle	93.2	99.52	81.97	79.52
Glass	86.46	93.33	74.77	70.8
Letter	96.22	99.98	96.88	68.133
Iris	99	1	1	1
breast cancer	62.7	65.8	73.38	71.78

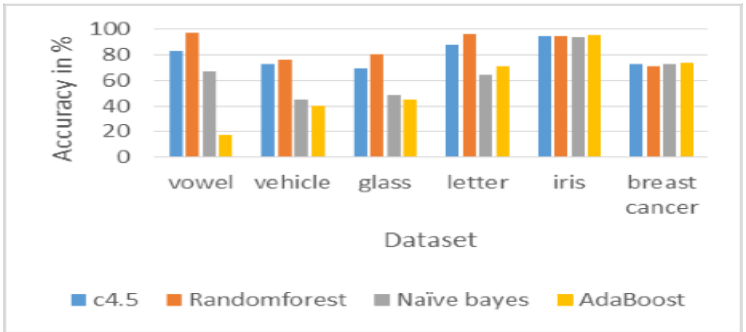

Figure I: Comparison for the values of the Accuracy of the compared algorithms.

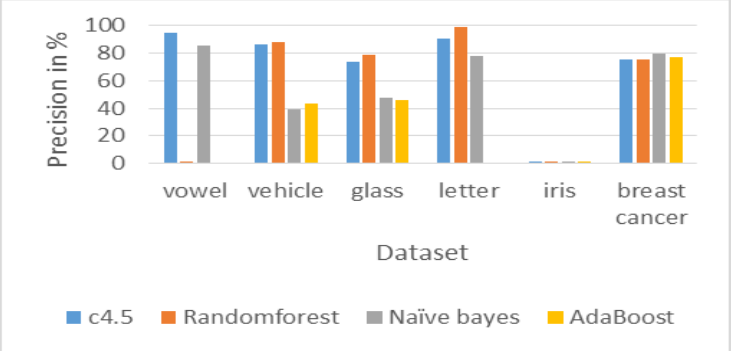

Figure II: Comparitive graph for Precision Percentage.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 4, Issue 5 , May 2017

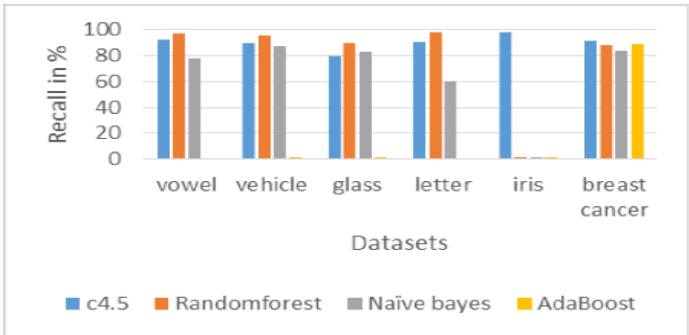

Figure III: Comparitive Graph for Recall% .

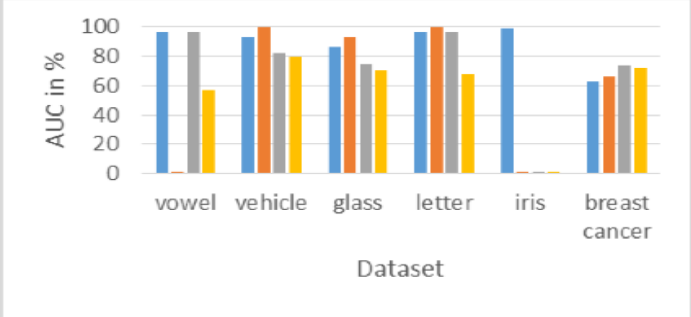

Figure IV: Comparative Graph for AUC%.

V. CONCLUSION

A new test-cost sensitive decision tree learning algorithm is proposed in this paper, which aims to keep the high classification accuracy meanwhile reduce the total test cost. Compared to C4.5,existing test-cost sensitive decision tree learning algorithms by adapting information theoretic measures to introduce the test cost degrade the classification accuracy when they reduce the total test cost, while our algorithm maintains the same classification accuracy as C4.5and at the same time significantly reduces the total test cost. This paper provides a new idea for research, i.e., it does not has to reduce the test costs at the cost of the loss of classification accuracy. We can reduce the total test cost and maintain the same classification accuracy as C4.5 simultaneously. For this purpose, a random attribute selection measure is presented. Instead of the greedy strategy, the proposed random attribute selection measure employs a random strategy to guide the selection of the optimal attribute for splitting. A random factor is introduced to tree building to make trees more biased in favor of the test cost or the classification accuracy. As already pointed out, there are two objectives in the task oftest-cost sensitive classification; one is decreasing the test cost, the other is improving the classification accuracy. Our current version transforms the test-cost sensitive classification problem into a constrained single-objective optimization problem. We believe that the use of more sophisticated multi-objective optimization methods could further improve the performance of the current algorithm and make its advantage stronger. This is a main direction for our future study. Besides, for simplicity, we assume that all features have discrete values only in this paper and thus all continuous features are discretized using a preprocessing step .However, in many real-world applications, continuous features are widespread and, therefore, extending it to directly handle applications with continuous features is another direction for our future study.

REFERENCES

[1.] J.R. Quinlan, J. Ross, Induction of decision trees, Mach. Learn. 1 (1) (1986)81–106.
[2.] J.R. Quinlan, C4.5: Programs for Machine Learning, San Mateo, MorganKaufmann, 1993.
[3.] L. Breiman, J. Friedman, J. Stone, Classification and Regression Trees, CRCPress, 1984.
[4.] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 4, Issue 5 , May 2017

[5.] C. Elkan, The foundations of cost-sensitive learning, in: International JointConference on Artificial Intelligence, Lawrence Erlbaum Associates, 2001, pp.973–978.
[6.] J.R. Quinlan, P.J. Ross, Inductive knowledge acquisition: a case study, in:Proceedings of the Second Australian Conference on Applications of ExpertSystems, Addison-Wesley Longman Publishing Co., 1987, pp. 137–156.
[7.] V. López, D. Rianõ, J.A. Bohada, Improving medical decision trees by combiningrelevant health-care criteria, Expert Syst. Appl. 39 (14) (2012) 11782–11791.
[8.] P.D. Turney, D. Peter, Cost-sensitive classification: empirical evaluation of ahybrid genetic decision tree induction algorithm, J. Artif. Intell. Res. (1995)369–409.
[9.] C.X. Ling, X. Charles, Decision trees with minimal costs, in: Proceedings of the21st International Conference on Machine learning, ACM, 2004.
[10.] Z. Qin, S. Zhang, C. Zhang, Cost-sensitive decision trees with multiple costscales., in: AI 2004: Advances in Artificial Intelligence, Springer BerlinHeidelberg, 2005, pp. 380–390.
[11.] T. Wang, Handling over-fitting in test cost-sensitive decision tree learning byfeature selection, smoothing and pruning, J. Syst. Softw. 83 (7) (2010)1137–1147.
[12.] F. Min, W. Zhu, A competition strategy to cost-sensitive decision trees.,in:Rough Sets andKnowledge Technology, Springer Berlin Heidelberg, 2012, pp.359–368.
[13.] Y. Weiss, Y. Elovici, L. Rokach, The cash algorithm-cost-sensitive attributeselection using histograms, Inf. Sci. 222 (2013) 247–268.
[14.] F. Min, H. He, Y. Qian, W. Zhu, Test-cost-sensitive attribute reduction, Inf. Sci.181 (22) (2011) 4928–4942.
[15.] F. Min, Q. Hu, W. Zhu, Feature selection with test cost constraint, Int. J.Approx. Reason.55 (1) (2014)167–179.
[16.] V. Bolón-Canedo, B.Remeseiro, N.Sánchez-Maroˇno, A.Alonso-Betanzos,Aframework for cost-based feature selection,Pattern Recognit. 47 (7) (2014)2481–2489.
[17.] V. Bolón-Canedo, B.Remeseiro,N.Sánchez-Maroˇno, A. Alonso-Betanzos,mC-ReliefF:an extension of ReliefF for cost-sensitive feature selection, in:Proceedings of International Conference of Agents and ArtificialIntelligence,Angers, France, 2014, pp. 42–51.
[18.] W.Qian, W.Shu, J.Yang, Y.Wang, Cost-sensitive feature selection onheterogeneous data, Adv. Knowl.Discov. Data Min.9078(2015) 397–408.
[19.] G.Kong, L.Jiang, C.Li, Beyond accuracy:learning selective Bayesian classifierswith minimal test cost, Pattern Recognit.Lett. 80 (2016) 65–171.
[20.] M. Núˇnez, Economic induction: a case study, in: Proceedings of the ThirdEuropean Working Session on Learning, Pitman Publishing, Glasgow, 1988,pp. 139–145.
[21.] S.W. Norton, Generating better decision trees, in: International JointConference on Artificial Intelligence, vol. 89, 1989, pp. 800–805.
[22.] M. Tan, J.C. Schlimmer, Two Case Studies in Cost-Sensitive ConceptAcquisition, 1990, pp. 854–860.
[23.] J.V. Davis, J. Ha, C.J. Rossbach, H.E. Ramadan, E. Witchel, Cost-sensitivedecision tree learning for forensic classification, in: Proceedings of the 17thEuropean Conference on Machine Learning, Springer Berlin Heidelberg, 2006,pp. 622–629.
[24.] A. Freitas, A. Costa-Pereira, P. Brazdil, Cost-sensitive decision trees applied tomedical data, in: Data Warehousing and Knowledge Discovery, SpringerBerlin Heidelberg, 2007, pp. 303–312.
[25.] A.Frank,A.Asuncion,UCI Machine Learning Repository,UniversityofCalifornia,School of Information and Computer Science,Irvine,CA, 2010.
[26.] L. Jiang, C. Li, Z. Cai, Learning decision tree for ranking, Knowl. Inf. Syst. 20(2009) 123–135.
[27.] L. Jiang, Learning random forests for ranking, Front. Comput. Sci. China 5 (1)(2011) 79–86.
[28.] X.Yang, Y.Qi, X. Song, J. Song, Test cost sensitive multigranulation rough set:model and minimal cost selection, Inf. Sci. 250 (2013) 184–199.
[29.] S. Sheng, C.X. Ling, Q. Yang, Simple test strategies for cost-sensitive decisiontrees., in: MachineLearning: ECML 2005, Springer Berlin Heidelberg, 2005,pp. 365–376.
[30.] J. Su, H. Zhang, A fast decision tree learning algorithm, in: Proceedings of theTwenty-First National Conference on Artificial Intelligence, AAAI Press, 2006,pp. 500–505.
[31.] Q. Yang, Test-cost sensitive classification on data with missing values, Knowl.Data Eng. 18 (5) (2006) 626–638.
[32.] L. Jiang, Random one-dependence estimators, Pattern Recognit. Lett. 32 (3)(2011) 532–539.
[33.] L. Jiang, Z. Cai, H. Zhang, Not so greedy: randomly selected naive Bayes, ExpertSyst. Appl. 39 (12) (2012) 11022–11028.
[34.] M. Hall, A decision tree-based attribute weighting filter for naive Bayes,Knowl. Based Syst. 20 (2) (2007) 120–126.
[35.] A. Ahmad, Decision tree ensembles based on kernel features, Appl. Intell. 41(3) (2014) 855–869.[36] I.H. Witten, E. Frank, M.A. Hall, Data Mining: Practical Machine Learning Toolsand Techniques, third ed., Morgan Kaufmann, 2011.

AUTHOR'S BIOGRAPHY

Sonia Jain, pursuing M.Tech (CSE) from Raipur Institute Of Technology (CSVTU), Bhilai.

Avinash Dhole, working as assistant professor since 2005 in subjects like theory of computation, compiler design, parallel processing and ANN.