



A postpruning decision algorithm based on loss minimization

Ahmed M. AHMED¹ , Ali Hakan ULUSOY^{2,*} , Ahmet RİZANER² 

¹Department of Computer Science, College of Computer Science, University of Bahri, Bahri, Sudan

²Department of Information Technology, School of Computing and Technology, Eastern Mediterranean University, Famagusta, North Cyprus via Mersin 10 Turkey

Received: 14.01.2019

Accepted/Published Online: 05.05.2019

Final Version: 26.07.2019

Abstract: In this paper, a post-pruning method known as zero-one loss function pruning (ZOLFP) that is based on zero-one loss function is introduced. The proposed ZOLFP method minimizes the expected loss, rather than evaluating the misclassification error rate of a node and its subtree. The subtree is pruned when expected loss of the node is less than or equal to the sum of the loss of its leaves. The experimental results demonstrate that ZOLFP method outperforms Un-pruned C4.5 Decision Tree (UDT-C4.5) algorithm, reduced error pruning (REP), and minimum error pruning (MEP) in terms of performance accuracy in all used datasets. It is also shown that the complexity of the proposed method ZOLFP is not more than the complexity of REP and MEP methods. Furthermore, the results show that ZOLFP method achieves satisfactory results compared to REP, MEP, and UDT-C4.5 algorithms in terms of precision score, recall score, true positive rate, false positive rate, F-measure, and area under ROC scores during the experiment process.

Key words: Zero-one, decision tree, loss minimization, pruning, loss function

1. Introduction

Due to the advance progress in information technology, a huge amount of data is generated. Different approaches have been introduced to analyze and handle this huge data for future decision making. Data mining is one of the newest powerful technologies used to extract valuable and meaningful knowledge from large databases. It is also known as knowledge discovery process or data pattern examination [1–3]. Several data mining techniques have been developed and used for knowledge extraction process from databases [4]. A decision tree that has a flow-like structure is one of the effective data mining techniques that is widely used due to its higher accuracy [5]. Induction is performed in top-down method and some attribute selection measures are used to select attributes such as information gain, gain ratio, and gain index. During the induction process, a decision tree may generate an unwanted and meaningless tree that is large in size with more complexity because of the overfitting problem. Therefore, pruning methods are introduced to combat overfitting problems that affect the efficiency and accuracy of data. Two types of pruning are used: postpruning and prepruning. In postpruning, initially the tree is grown and then unsuitable branches are removed. Hence, postpruning decision tree produces the (complete) tree and then adjusts it in order to improve the classification accuracy on unseen instances. In prepruning, the overfitting is checked during the tree building process. There are two techniques for prepruning: minimum number of object pruning and chi-square pruning.

Several postpruning methods have been introduced in the literature such as reduced error pruning (REP),

*Correspondence: alihakan.ulusoy@emu.edu.tr

error complexity pruning, minimum error pruning (MEP), and cost-based pruning. Among these techniques, REP produces smaller trees with better accuracy. REP traverses the decision tree in postorder by checking all internal nodes and replaces subtrees with their leaf nodes when the former misclassification error rate is no smaller than the later. In this method as in some other postpruning methods, the dataset is divided into three sets as training set that is used for training, the validation set that is used for pruning the tree, and test set that is used for an unbiased estimation over future unseen instances [6]. One of the advantages of this method is its linear computational complexity, but when the test set is much smaller than the training set, it tends to overprune.

The aim of most pruning algorithms is to minimize the expected error rate of the decision tree. Recently, some postpruning methods have been introduced to prune the tree with respect to the loss matrix [7–9]. Thus, in such cases, it may be desirable to prune the tree by estimating the expected loss instead of estimating the misclassification error rate. Pruning for loss minimization is another point of view in pruning that may result in different pruning behavior than does pruning for error minimization. This type of pruning relies on the probability distribution to adjust the prediction to minimize the expected loss or to supply a confidence level associated with the prediction [10]. While some researchers employed statistical methods to tackle the overfitting problem [11], some researchers adopted the idea of incorporating the general loss matrix into C4.5 decision tree algorithm to reduce the misclassification cost or to minimize the misclassification loss. These algorithms developed heuristic splitting methods that perform dual tasks. The first task is to select the splitting that minimizes the misclassification cost and the second task is to select proper split for subsequent splits.

In this paper, a new decision tree postpruning method known as zero-one loss function pruning (ZOLFP) that is based on zero-one loss function is introduced for loss minimization. Instead of estimating the misclassification error rate, the loss (or cost) of misclassification is estimated for each node and its subtree. If the loss of the node is less than or equal to the sum of the loss for the whole subtree under that node, the subtree is pruned. Several experiments are conducted to investigate the effectiveness of proposed ZOLFP method and its results are compared with those of the REP and MEP methods and Un-pruned C4.5 Decision Tree (UDT-C4.5).

The paper is organized as follows. Section 1 presents the introduction and research motivation, whereas, in Section 2, the literature review and related works are discussed. The proposed loss minimization based on ZOLFP algorithm is investigated and studied with a running example in Section 3. In Section 4, the experimental results about the proposed method are presented and compared with some existing approaches. Finally, the conclusion of the paper and future work are explained in Section 5.

2. Related Works

Over the past decades, different postpruning techniques have been introduced by researchers to investigate the overfitting problem. In [4], the researchers performed a comparative study to investigate and analyze six well-known postpruning techniques, namely, REP, pessimistic error pruning, MEP, critical value pruning, cost complexity pruning, and error-based pruning. The researchers highlighted the theoretical weakness and strengths of each method. Their results showed that REP produced the smallest subtree with the lowest error rate with respect to the pruning set. However, in [12], the researchers introduced a new decision tree algorithm based on J48 and REP. The new method was compared to original J48 decision tree and the results showed that their method produced a smaller tree and produced better performance accuracy. In [13], the researchers implemented decision tree induction algorithm with REP technique to improve the performance accuracy. Their proposed method generated an optimal decision tree with less complexity and better performance accuracy.

The research in [14] addressed the problem of multilabel classification where each example can belong to more than one class at the same time. The researchers introduced a new pruning technique known as PruDent. In terms of performance accuracy, PruDent was more accurate when compared with other state-of-the-art approaches and its computational costs were linear. On the other hand, one of the drawbacks of this method was the reliance on confidence scores. Furthermore, authors in [15] adopted ability, stability, and scale as new classification standards for classification evaluation. Results showed that the improved method solved problems better than single standard evaluation methods and reflected more advantages. Thus, this improved method produced a more balanced classification performance and less model complexity. The researchers in [16] introduced a new decision tree algorithm that was named as competition cost-sensitive C4.5 for numeric data based on C4.5. They designed a heuristic function that was based on the test cost and the information gain ratio. The results showed that the proposed postpruning algorithm was more effective and stable, and it produced a decision tree with a lower cost.

In [17], the researchers developed a postpruning decision tree algorithm that was based on the Bayesian theory, where the branches that were generated by the C4.5 algorithm validated by the Bayesian theorem. This method produced a small decision tree since the branches out of the condition range were removed. The results showed that the proposed method produced a simpler decision tree compared to the original C4.5 algorithm. In [6], various measurement techniques were utilized to evaluate the performance of postpruning methods like accuracy, stability, and simplicity. A multiobjective evaluation was proposed to select the best subtree during the postpruning process. Moreover, the researchers developed a procedure for obtaining the optimal subtree based on user-provided preference and value function information. The researchers in [18] conducted a comparison study for REP method in decision tree. The performance of REP was analyzed and they deduced that if the algorithm produced a simple tree with low accuracy it meant that the algorithm computed high misclassification during instance learning process. The results showed that J48 and REP produced a tree with high accuracy of classification with less complexity. The researchers in [19] introduced a new pruning method that aimed to improve both classification accuracy and tree size. The newly introduced method was first applied to decision tree by implementing prepruning at the inducing phase of the tree, and postpruning after the inducing process. The researchers in [20] introduced a new decision tree pruning method based on backpropagation neural networks, called soft-pruning. Firstly, C4.5 method was employed for decision tree induction and then the obtained trees were pruned by using the soft-pruning method. The experiment results indicated that the soft-pruning method performed better than original C4.5 pruned and unpruned trees.

The researchers in [21] proposed a new C5.0 classifier method that employed feature selection, cross validation, REP, and model complexity for the original C5.0 method to reduce the tree size and improve the accuracy. The experimental results showed that when feature selection was applied, the attribute space was reduced for feature set while applying the cross validation technique provided a more reliable estimate of predictive. On the other hand, when the model complexity was increased, the accuracy of the classification was also increased, and whenever the REP technique was applied, the overfitting rate of the decision tree was reduced. The accuracy of the proposed method was improved compared to the original C5.0 method. A new pruning method called multilevel pruned classifier that integrated the pruning phase into the building phase was developed in [22]. The experiments were conducted on the dataset with and without pruning in terms of complexity and classification accuracy.

The researchers in [23] proposed a decision tree method that adopted rough set theory for pruning. The proposed method introduced depth-fitting ratio which involved both the depth and the explicit degrees of the

subtrees under evaluation. The experimental results demonstrated that the newly proposed method was feasible and effective for pruning. The constructed decision tree sizes were quite reduced, while the prediction accuracy was well improved. Furthermore, in [24], a multistrategy pruning algorithm was introduced to trim the tree. During the pruning process, three groups of strategies were implemented to get the optimal solution, which were namely, simple size, degree of matching, and scale of the tree. The experimental results illustrated that the multistrategy pruning algorithms for decision tree pruning improved the efficiency and accuracy of intrusion detection system. The transactions were classified into four risk levels instead of classifying it either fraud or nonfraud, such that the proposed method showed a promising progress.

In [25], the researchers introduced a decision tree pruning method based on genetic algorithms. The experimental results indicated that applying genetic algorithms in decision tree pruning was effective and feasible, and the pruning operation was converted to optimize the edge weight. The new approach was compared with some decision tree pruning techniques including cost-complexity pruning, pessimistic error pruning, and REP. The results showed that the new method had better or equal effect with other pruning methods. A unifying framework based on the four-tuple (space, operators, evaluation function, and search strategy) was introduced in [26]. The pruning methods were investigated by means of this framework and their common aspects, strengths, and weaknesses were described. The experiments were conducted by utilizing these six well-known pruning methods. A comparison was performed between the mentioned pruning methods, and their common aspects, strengths, and weaknesses were also investigated. The results demonstrated the fact that pruning methods did not reduce the productivity, but they could enhance the final tree accuracy.

3. Proposed zero-one loss function pruning algorithm based on loss minimization

In the past decade, many algorithms have been introduced to minimize the misclassification error rate, while some researches attempted to minimize the expected loss or the cost of misclassifying an example in the dataset. Although, recent researches continued to develop more accurate algorithms to enhance the misclassification error rate, applications in business, medicine, and science have shown that real problems require more subtle measures of performance [27–30]. Furthermore, in general, various kinds of errors have different costs. The performance of classifiers is evaluated by estimating the error rate obtained from the dataset which represents the incorrect classified examples. It is sometimes suitable to evaluate classifiers by considering the cost of the misclassification errors based on loss matrix.

Several researchers have adopted the idea of incorporating the general loss matrix into C4.5 decision tree algorithm to reduce the misclassification cost or to minimize the misclassification loss [31–34]. These algorithms introduced heuristic splitting methods that perform dual tasks. Initially, they specify the splitting that minimizes the loss and secondly, they select the proper split for subsequent splits. Normally, the loss matrices are T_c by T_c matrices where T_c is the number of classes, while the rows represent the predicted classes by the classifier algorithm and the columns represent the correct classes. The loss function $L(i | j)$ gives the “loss” of predicting class i when the true class is j , and diagonal elements, $L(i | i)$, are always zero [35]. In general, loss functions state exactly how each action costs, for instance, the loss function $L(a_i | C_j)$ indicates the loss obtained for taking action a_i when the class is C_j .

In this paper, a new decision tree postpruning method that is based on loss minimization is introduced. This method focuses on employing the Zero-One loss function in C4.5 decision tree to avoid the over-fitting problem and improve the performance. The expected loss or so-called the conditional risk of Zero-One loss function is defined by:

$$R_{0/1}(a_i | x) = \sum_{j=1, j \neq i}^{T_c} P(C_j | x), \tag{1}$$

where $P(C_j | x)$ is the probability that an example x belongs to a specific class C_j .

Here, Laplace method is utilized to estimate the probability that an example belongs to class i in the specific node. If there are N_i examples of class i at a node and T_c classes, then the probability P_{ic} that an example at this node belongs to class i is estimated by [36]:

$$P_{ic} = (N_i + 1)/(T_c + N_T), \tag{2}$$

where N_T is the total number of examples at that node.

Thus, when Laplace method is employed in (1) to estimate the probability that an example belongs to specific class in a node, the zero-one loss is computed as follows:

$$R_{0/1} = 1 - P_{ic}. \tag{3}$$

The proposed pruning method prunes the decision tree in the bottom-up fashion by adopting REP approach. However, instead of estimating the misclassification error rate, the expected loss is estimated by applying zero-one loss function method. For that reason, the proposed method is named as zero-one loss function pruning. The zero-one loss for each node is computed by applying (3). Then, the expected loss, L_{ex} , for each class is computed by multiplying the zero-one loss by the loss function related to that node as:

$$L_{ex} = R_{0/1}L(i | j) = (1 - P_{ic})L(i | j). \tag{4}$$

The pruning process is performed by dividing the dataset into two sets as training set and pruning set. The training set is used to create the decision tree and then it is tested by using the pruning set. During the pruning, the expected loss for each tested node (L_{ex}^p) is computed as the sum of the expected losses of its classes. If the expected loss for the tested node is less than or equal to the sum of the expected loss for the whole subtree under that tested node (L_{ex}^t), these tested nodes are converted to leaf nodes, and leaf nodes are labeled by the least expected loss class that is equivalent to label of the class with the majority class label in decision tree nodes. The converted test nodes to leaf nodes are given the class label of the least expected loss class. The tree is created based on C4.5 decision tree algorithm and then the pruning process is performed by using the newly introduced ZOLFP method that is shown in Figure 1.

How the proposed method works is shown with a running example in Figure 2. In this example, there are two classes for unpruned decision tree, a person is classified as healthy or sick accordingly. The proposed method, ZOLFP, is examined to prune this tree based on loss (cost) approach. The ZOLFP method computes the expected loss for each node by applying Eq. (4). The loss function for each node is indicated where h denotes the healthy class and s denotes the sick class. In Figure 2a, the subtree should not be pruned when ZOLFP method is applied because the parent loss is greater than the sum of losses of the children (10.4 versus 5.9). Figure 2b shows the reverse situation. Here, the ZOLFP method prunes the tree because the parent has lower loss than the sum of losses of the children (9.2 versus 14.2). Thus, the subtree is pruned.

4. Experimental result of the proposed ZOLFP method

In this section, the performance of the proposed ZOLFP method is demonstrated and compared to UDT-C4.5, REP, and MEP approach on different datasets.

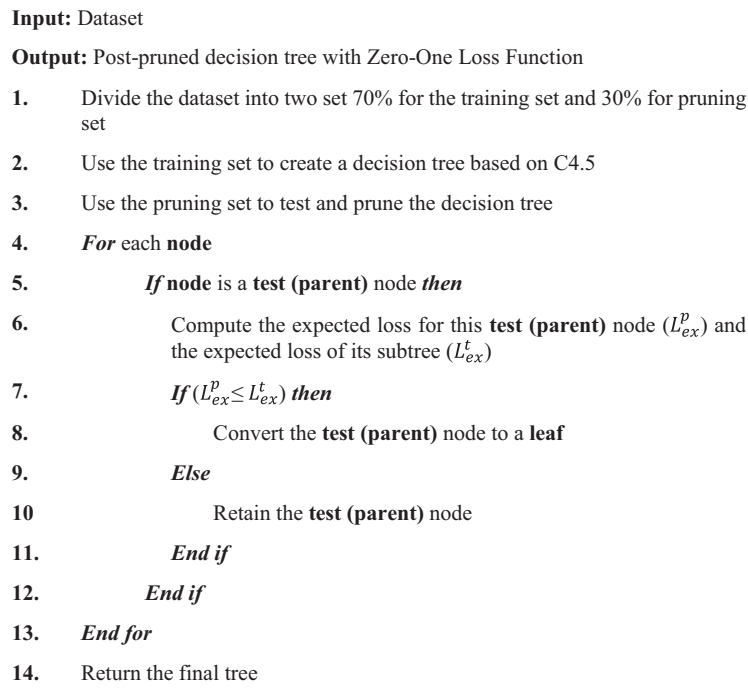


Figure 1. The algorithm of ZOLFP.

4.1. Experiment data

The experiment datasets are collected from UCI machine learning repository (<https://archive.ics.uci.edu/ml>). In this paper, six different datasets have been used to evaluate the proposed method. Table 1 presents the number of instances, the number of classes, and the number of attributes for the datasets.

Table 1. Dataset description.

Datasets	Number of instances	Number of attributes	Number of classes
Car	1728	7	4
Diabetes	768	8	2
Labour	57	16	2
Blogger	100	6	2
User knowledge	258	6	4
Flare solar	1066	13	3

4.2. Experiment process

The experiment process is carried out by using java eclipse combined with Weka. Additionally, Weka attribute evaluator techniques namely One Rule (OneR) and Information Gain (InfoGainAttributeEval) are employed to select the attributes with high impact on the datasets and remove the worst attributes that affect the accuracy performance as shown in Table 2.

The datasets in Table 1 are trained and tested with 10-fold cross validation by the proposed method ZOLFP after the worst attributes are removed by OneR attribute evaluator and InfoGainAttributeEval attribute

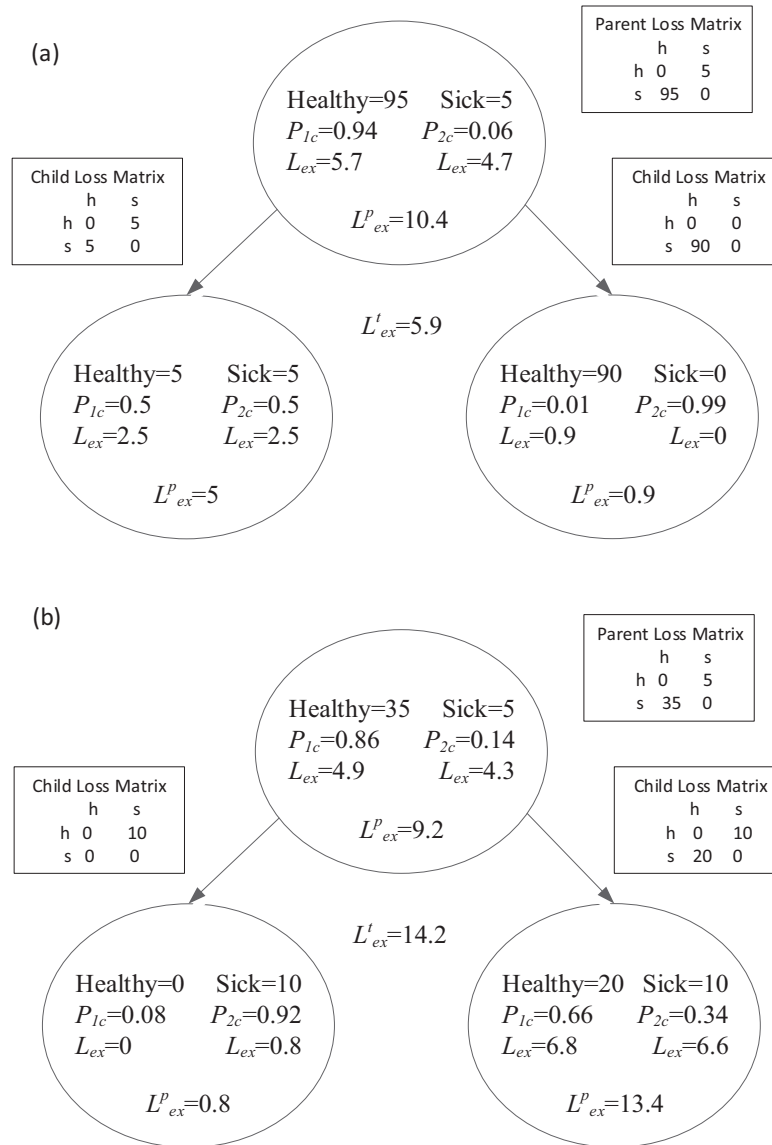


Figure 2. A running example of the ZOLFP method.

evaluator. A comparison is conducted between the 10-fold attribute evaluator's result and the 60% hold-out validation results as indicated in Table 3. The results in Table 3 show that 60% hold-out validation method obtains better performance for ZOLFP method compared to the case when 10-fold cross-validation is used.

It is also obvious that ZOLFP with 10-fold OneR attribute evaluator achieves better accuracies compared to ZOLFP with 10-fold InfoGainAttributeEval evaluator with four scores to the former and none scores for the latter, whereas both methods have the same scores in one dataset. Because hold-out method suffers from some limitation, it is not sufficient to give a reliable model so that cross-validation methods are considered in this paper.

OneR attribute evaluator with 10-fold cross validation method is employed in this experiment since it shows better performance compared to InfoGainAttributeEval attribute evaluator with 10-fold cross validation method.

Table 2. Indicates the removed attributes by OneR and InfoGainAttributeEval attribute evaluators.

Datasets	OneR	InfoGainAttributeEval
Car	perspons, buying	lug-boot, doors
Diabetes	mass, pedi, skin	pres, pedi, skin
Labour	wage2.wage, shift_diff, dur, hours.hrs, wage3.wage	education-allowance, standby-pay, working-hours, duration
Blogger	lmt, lpss	lmt, lpss
User knowledge	STG, SCG	STR, SCG
Flare solar	previous 24 hours activity code, historically, M-class production, area of the largest spot, area, did region become historically	historically, C-class production, evolution, did region become historically, area of the largest spot

Table 3. Performance accuracy results for ZOLFP with OneR and InfoGainAttributeEval attribute evaluators.

Datasets	Accuracy (%)		
	60% training, 40% testing, OneR (Hold out)	10-fold, OneR	10-fold, InfoGainAttributeEval
Car	97	95	84
Diabetes	89	76	73
Labour	95	82	77
Blogger	87	81	77
User knowledge	93	84	75
Flare solar	99	99	99

4.3. Experiment result analysis

Typically, the complexity of the tree is measured by one of the metrics such as total number of nodes (tree size), total number of leaves, tree depth, and the number of attributes used [18, 37]. Table 4 shows the accuracy and the tree complexity in terms of tree size as the total number of nodes and leaves for ZOLFP, REP, and MEP approaches. The results are also compared with UDT-C4.5 algorithm to illustrate the effect of pruning. For all the datasets, the proposed ZOLFP method produces better accuracy compared to REP and MEP methods. In terms of complexity, the proposed method, ZOLFP, produces smaller tree size than REP and MEP for three datasets. For the dataset which the ZOLFP complexity is higher, the accuracy of ZOLFP is better than the accuracies of REP and MEP. On the other side, it is obvious that the proposed method, ZOLFP, produces a smaller tree with better accuracy compared to UDT-C4.5 for all datasets.

Additionally, precision score, recall score, the weighted averages of true positive (TP) and false positive (FP) rates, F-measure, and area under receiver operating characteristic (ROC) are also investigated to evaluate the performance of ZOLFP, REP, and MEP methods as shown in Table 5. For precision scores, ZOLFP produces better scores than REP and MEP in all datasets, where in terms of recall scores, ZOLFP produces better scores in five datasets. For Diabetes dataset, the recall scores of ZOLFP and REP are the same but the precision score of ZOLFP is better than REP’s for the same dataset.

The proposed method produces the highest TP rate and F-measure in all datasets and the lowest FP rates for four datasets. Consequently, the proposed ZOLFP method produces highest scores in terms of the area under ROC and compared to REP and MEP with four scores.

Table 4. Performance and tree size for ZOLFP, REP, and MEP methods.

Datasets	Algorithms	Accuracy (%)	Number of Nodes	Number of Leaves
Car	UDT-C4.5	94	186	134
	ZOLFP	95	88	87
	REP	94	182	129
	MEP	88	113	81
Diabetes	UDT-C4.5	73	43	22
	ZOLFP	76	31	16
	REP	74	41	21
	MEP	74	15	8
Labour	UDT-C4.5	72	47	39
	ZOLFP	77	36	30
	REP	65	13	8
	MEP	75	18	17
Blogger	UDT-C4.5	73	43	28
	ZOLFP	77	33	23
	REP	72	15	13
	MEP	75	12	9
User knowledge	UDT-C4.5	75	404	399
	ZOLFP	77	402	399
	REP	75	81	80
	MEP	74	81	80
Flare solar	UDT-C4.5	95	25	21
	ZOLFP	99	7	6
	REP	97	19	16
	MEP	96	25	21

5. Conclusions

In this paper, a decision tree pruning method known as ZOLFP that takes into accounts both classification accuracy and tree size is introduced. We proposed a zero-one loss function pruning method that minimizes the expected loss of the produced tree. This method is based on loss minimization, instead of estimating the misclassification error rate the loss of misclassification is estimated. The proposed method is trained on six different datasets and a comparison is conducted between the proposed method ZOLFP, UDT-C4.5, REP, and MEP methods in terms of performance accuracy and complexity.

The experiment results show that the ZOLFP method outperforms REP and MEP methods in term of performance accuracy for all datasets employed. While in terms of complexity, the ZOLFP method produces less tree size than REP and MEP in three datasets. On the other hand, the ZOLFP method outperforms UDT-C4.5 method in terms of accuracy and complexity for all datasets. Furthermore, the results also show that the proposed method, ZOLFP, yields reasonable performance scores for precision score, recall score, TP rate, FP rate, and area under ROC compared to REP and MEP methods.

The proposed algorithm adopts a postpruning bottom-up method for C4.5 decision tree algorithm. As

Table 5. Precision score, recall score, TP rate, FP rate, F-measure, and area under ROC comparisons

Datasets	Algorithms	Precision (%)	Recall (%)	TP rate (%)	FP rate (%)	F-measure (%)	Area under ROC (%)
Car	UDT-C4.5	94	94	94	3	94	96
	ZOLFP	95	95	95	2	95	98
	REP	94	94	94	3	94	98
	MEP	89	88	88	6	88	96
Diabetes	UDT-C4.5	72	73	73	31	72	74
	ZOLFP	76	76	76	32	76	83
	REP	75	74	74	29	75	81
	MEP	74	74	74	31	74	77
Labour	UDT-C4.5	72	72	72	36	72	72
	ZOLFP	76	77	77	27	81	84
	REP	65	65	65	29	78	82
	MEP	74	75	75	31	77	79
Blogger	UDT-C4.5	72	73	73	39	72	73
	ZOLFP	76	77	77	33	80	82
	REP	70	72	72	36	78	81
	MEP	74	75	75	36	76	73
User knowledge	UDT-C4.5	76	75	76	10	95	89
	ZOLFP	77	77	77	6	99	95
	REP	75	75	75	7	97	94
	MEP	75	74	75	11	96	91
Flare solar	UDT-C4.5	95	95	95	9	64	54
	ZOLFP	97	99	99	8	75	64
	REP	97	97	97	8	66	72
	MEP	96	96	96	9	57	31

future works, the proposed method ZOLFP can be applied on C5.0 decision tree classifier and can be also modified for other decision tree based classifiers such as best first tree and random forest [38]. Additionally, the running time and complexity analysis of the algorithms can be investigated.

Acknowledgment

We are thankful to the editor and anonymous referees for their constructive suggestions and comments.

References

- [1] Anyanwu MN, Shiva SG. Comparative analysis of serial decision tree classification algorithms. *International Journal of Computer Science and Security* 2009; 3 (3): 230-240.
- [2] Han DH, Zhang X, Wang GR. Classifying uncertain and evolving data streams with distributed extreme learning machine. *Journal of Computer Science and Technology* 2015; 30 (4): 874-887. doi: 10.1007/s11390-015-1566-6

- [3] Sumathi S, Sivanandam S. Data mining tasks, techniques, and applications. In: *Introduction to Data Mining and its Applications*. New York, NY/Berlin: Springer, 2006. pp. 195-216.
- [4] Esposito F, Malerba D, Semeraro G, Kay J. A comparative analysis of methods for pruning decision trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1997; 19 (5): 476-491.
- [5] Han J, Pei J, Kamber M. *Data mining: concepts and techniques*. 3rd ed. Waltham, MA, USA: Morgan Kaufmann, Elsevier, 2011.
- [6] Osei-Bryson KM. Post-pruning in decision tree induction using multiple performance measures. *Computers & Operations Research* 2007; 34 (11): 3331-3345. doi: 10.1016/j.cor.2005.12.009
- [7] Wang T, Qin Z, Jin Z, Zhang S. Handling over-fitting in test cost-sensitive decision tree learning by feature selection, smoothing and pruning. *Journal of Systems and Software* 2010; 83 (7): 1137-1147. doi: 10.1016/j.jss.2010.01.002
- [8] Esmeir S, Markovitch S. Anytime learning of anycost classifiers. *Machine Learning* 2011; 82 (3): 445-473. doi: 10.1007/s10994-010-5228-1
- [9] Sheng VS, Ling CX. Cost-sensitive learning. In: *Encyclopedia of Data Warehousing and Mining*. 2nd ed. New York, NY, USA: IGI Global, 2009. pp. 339-345.
- [10] Bradford JP, Kunz C, Kohavi R, Brunk C, Brodley CE. Pruning decision trees with misclassification costs. In: *10th European Conference on Machine Learning (ECML 98)*; Chemnitz, Germany; 1998. pp. 131-136.
- [11] Katz G, Shabtai A, Rokach L, Ofek N. ConfDtree: a statistical method for improving decision trees. *Journal of Computer Science and Technology* 2014; 29 (3): 392-407. doi: 10.1007/s11390-014-1438-5
- [12] Kapoor P, Rani R. Efficient decision tree algorithm using J48 and reduced error pruning. *International Journal of Engineering Research and General Science* 2015; 3 (3): 1613-1621.
- [13] Patel RR, Aluvalu R. A reduced error pruning technique for improving accuracy of decision tree learning. *International Journal of Engineering and Advanced Technology* 2014; 3 (5): 8-11.
- [14] Alali A, Kubat M. Prudent: A pruned and confident stacking approach for multilabel classification. *IEEE Transactions on Knowledge and Data Engineering* 2015; 27 (9): 2480-2493. doi: 10.1109/TKDE.2015.2416731
- [15] Xie H, Shang F. The study of methods for post-pruning decision trees based on comprehensive evaluation standard. In: *IEEE 2014 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 14)*; Xiamen, China; 2014. pp. 903-908.
- [16] Xu ZL, Min F, Zhu W. Cost-sensitive decision trees with post-pruning and competition for numeric data. *Journal of Computational Information Systems* 2013; 9 (20): 7991-8001.
- [17] Zhang W, Li Y. A post-pruning decision tree algorithm based on Bayesian. In: *International Conference on Computational and Information Sciences (ICCIS 13)*; Shiyang, Hubai, China; 2013. pp. 988-991.
- [18] Mohamed WNH, Salleh MNM, Omar AH. A comparative study of Reduced Error Pruning method in decision tree algorithms. In: *IEEE 2012 International Conference on Control System, Computing and Engineering (ICCSCE 12)*; Penang, Malaysia; 2012. pp. 392-397.
- [19] Mahmood AM, Mrithyumjaya PGVGK, Kuppa R. A new pruning approach for better and compact decision trees. *International Journal on Computer Science & Engineering* 2010; 2 (8): 2551-2558.
- [20] Kijirikul B, Chongkasemwongse K. Decision tree pruning using backpropagation neural networks. In: *IEEE 2001 International Joint Conference on Neural Networks (IJCNN 01)*; Washington, USA; 2001. pp. 1876-1880.
- [21] Galathiya A, Ganatra A, Bhensdadia C. Improved decision tree induction algorithm with feature selection, cross validation, model complexity and reduced error pruning. *International Journal of Computer Science and Information Technologies* 2012; 3 (2): 3427-3431.
- [22] Patil DD, Wadhai V, Gokhale J. Evaluation of decision tree pruning algorithms for complexity and classification accuracy. *International Journal of Computer Applications* 2010; 11 (2): 23-30.

- [23] Hall MP. Remote detection and predicted locations of NIPF fuel treatments in central Oregon. MS, Oregon State University, Oregon, USA, 2015.
- [24] Wang H, Chen B. Intrusion detection system based on multi-strategy pruning algorithm of the decision tree. In: IEEE 2013 International Conference on Grey Systems and Intelligent Services (GSIS 13); Macao; 2013. pp. 445-447.
- [25] Chen J, Wang X, Zhai J. Pruning decision tree using genetic algorithms. In: IEEE 2009 International Conference on Artificial Intelligence and Computational Intelligence (AICI 09); Beijing, China; 2009. pp. 244-248.
- [26] Esposito F, Malerba D, Semeraro G, Tamma V. The effects of pruning methods on the predictive accuracy of induced decision trees. *Applied Stochastic Models in Business and Industry* 1999; 15 (4): 277-299.
- [27] Zhou A, Qian W, Qian H, Jin W. A new classification method to overcome over-branching. *Journal of Computer Science and Technology* 2002; 17 (1): 18-27.
- [28] Rahmati O, Kornejady A, Samadi M, Deo RC, Conoscenti C et al. PMT: New analytical framework for automated evaluation of geoenvironmental modelling approaches. *Science of the Total Environment* 2019; 664: doi: 10.1016/j.scitotenv.2019.02.017
- [29] Goerigk S, Hilbert S, Jobst A, Falkai P, Buhner M et al. Predicting instructed simulation and dissimulation when screening for depressive symptoms. *European Archives of Psychiatry and Clinical Neuroscience* 2018; 1-16. doi: 10.1007/s00406-018-0967-2
- [30] Bahnsen AC, Stojanovic A, Aouada D, Ottersten B. Cost sensitive credit card fraud detection using Bayes minimum risk. In: IEEE 2013 12th International Conference on Machine Learning and Applications (ICMLA 13); Florida, USA; 2013. pp. 333-338.
- [31] Fawcett T, Provost F. Adaptive fraud detection. *Data Mining and Knowledge Discovery* 1997; 1 (3): 291-316.
- [32] Thomas P, El Haouzi HB, Suhner MC, Thomas A, Zimmermann E et al. Using a classifier ensemble for proactive quality monitoring and control: the impact of the choice of classifiers types, selection criterion, and fusion process. *Computers in Industry* 2018; 99: 193-204. doi: 10.1016/j.compind.2018.03.038
- [33] Wu G, Tian Y, Liu D. Cost-sensitive multi-label learning with positive and negative label pairwise correlations. *Neural Networks* 2018; 108: 411-423. doi: 10.1016/j.neunet.2018.09.003
- [34] Kubat M, Holte R, Matwin S. Learning when negative examples abound. In: 9th European Conference on Machine Learning (ECML 97); Prague, Czech Republic; 1997. pp. 146-153.
- [35] Loh WY. Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2011; 1 (1): 14-23.
- [36] Pazzani M, Merz C, Murphy P, Ali K, Hume T et al. Reducing misclassification costs. In: 11th International Conference on Machine Learning (ICML 94); New Brunswick, NJ, USA; 1994. pp. 217-225.
- [37] Kotsiantis SB. Decision trees: A recent overview. *Artificial Intelligence Review* 2013; 39 (4): 261-283. doi: 10.1007/s10462-011-9272-4
- [38] Chen W, Zhang S, Li R, Shahabi H. Performance evaluation of the gis-based data mining techniques of best-first decision tree, random forest, and Naive Bayes tree for landslide susceptibility modeling. *Science of the Total Environment* 2018; 644: 1006-1018. doi: 10.1016/j.scitotenv.2018.06.389