[Future Computing and Informatics Journal](https://digitalcommons.aaru.edu.jo/fcij)

Volume 3
Issue 2 *Vol 3,Issue 2,2018*

Article 12

2018

Non-sequential partitioning approaches to decision tree classifier

Shankru Guggari
Dept. of Computer Science and Engineering, B.M.S. College of Engineering, shankar286@gmail.com

Vijayakumar Kadappa
Dept. of Computer Applications, B.M.S. College of Engineering, Bengaluru, vijaykumark.mca@bmsce.ac.in

V. Umadevi
Dept. of Computer Science and Engineering, B.M.S. College of Engineering, umadevi.cse@bmsce.ac.in

Available online at www.sciencedirect.com

ScienceDirect

Future Computing and Informatics Journal 3 (2018) 275–285
http://www.journals.elsevier.com/future-computing-and-informatics-journal/

Non-sequential partitioning approaches to decision tree classifier

Shankru Guggari [a,*], Vijayakumar Kadappa [b], V. Umadevi [a]

[a] Dept. of Computer Science and Engineering, B.M.S. College of Engineering, Bengaluru, 560019, India
[b] Dept. of Computer Applications, B.M.S. College of Engineering, Bengaluru, 560019, India

Received 29 March 2018; accepted 21 June 2018
Available online 18 September 2018

Abstract

Decision tree is a well-known classifier which is widely used in real-world applications. It is easy to interpret, however it suffers from instability and lower classification performance for high-dimensionality datasets due to curse of dimensionality. Feature set partitioning is a novel concept to address the higher dimensionality problem by dividing the feature set into subsets (blocks). Many of the existing partitioning based decision tree approaches are sequential in nature, which lack logical relationships amongst the features. In this work, we propose novel non-sequential feature set partitioning methods by exploiting the ideas of Ferrer Diagram and Bell Triangle to create feature blocks with a mix of low, medium, and high correlation features. The experimental results on 11 UCI and KEEL datasets demonstrate the superiority of the proposed partitioning methods, upto 5% higher classification accuracy, over NBTree, BFTree, Serial-CMFP partitioning method, and classical decision tree techniques. The proposed methods also exhibit improved stability as compared to other decision tree methods.

Keywords: Decision tree; Correlation; Ferrer diagram; Bell triangle; Partitioning

1. Introduction

Decision tree (DT) is one of the most popular classification models in machine learning and data mining. DT classifier is also known as hierarchical classifier and widely applied for supervised classification in data analytics. The main aim of the DT classifier is to build a model that predicts the target variable based on various input variables. It can handle both continuous and categorical values. DTs are successfully used in various applications like Finding human location [1], Prediction of student success rate [2], Money laundering risk evaluation [3], Security assessment in power system [4], Forecasting copper prices [5], Hard drive failure prediction [6], etc. DTs suffer from several challenges like Instability (that is, slight changes in training samples lead to significant changes in the resultant tree) and classification generalization ability for higher dimensional datasets to name a few. Classical decision trees like CART [7], C4.5 [8], C5.0 [9], NBTree [10], BFTree [11] suffers from these challenges.

The partitioning based methods in pattern recognition are widespread and known to be efficient as compared to traditional methods by utilizing local information [12–16]. There are two types of Partitioning methods: (i) Horizontal partitioning – the set of data instances is divided into subset of data instances (mini databases), and (ii) Vertical partitioning - the feature set is divided into feature subsets for every data instance [12]. In this work, we investigate on Vertical partitioning (also known as Feature set partitioning) methods to improve the classification ability and stability of a Decision tree classifier. In this direction, Seetha and Murthy [13] introduce serial vertical partitioning technique (Serial-CMFP) where in each partition has equal number of features, SVM and KNN classifiers are applied on each partition, and classifier decisions are combined using majority vote rule. Many existing vertical approaches [13] divide the feature set in sequential fashion which is not logical as they do not

* Corresponding author.
E-mail addresses: shankar286@gmail.com (S. Guggari), vijaykumark.mca@bmsce.ac.in (V. Kadappa), umadevi.cse@bmsce.ac.in (V. Umadevi).

Peer review under responsibility of Faculty of Computers and Information Technology, Future University in Egypt.

https://doi.org/10.1016/j.fcij.2018.06.003

consider inter relationships among the features. In this work, we propose two novel vertical partitioning approaches based on the ideas from Ferrer Diagram [17] and Bell Triangle [18]. Ferrer Diagram is used to find scientist's publication output using a bibliometric tool called citation triad [19] and others use it to represent the scientific output by constructing a new indicator, h-index [20]. Bell triangle is applied in various real world scenarios and shows connection between graph composition and Bell triangle [18,21]. We exploit the ideas of Ferrer Diagram [17] and Bell Triangle [18] to create novel feature blocks of a given dataset. The proposed methods are proved to be superior in terms of classification and stability over other decision tree methods.

The paper is organized as follows: Section 2 describes a review of Decision tree methods. The related work is presented in Section 3. We present our proposed work formally in Section 4. Results and Analysis are presented in Section 5 & Section 6 and we conclude in Section 7.

2. Review of decision tree methods

In this section, we briefly review the decision trees to aid the readers for easier understanding of other sections.

2.1. Classification and regression tree (CART)

CART constructs binary tree and splits dataset based on gini index. It prunes the tree using minimal cost complexity, which is computed using number of leaves and percentage of data instances mis-classified by the tree [7]. CART builds the regression trees and predicts the class label based on the weighted mean of the node [22].

2.2. C4.5

C4.5 is a popular algorithm and an extension of basic ID3 algorithm. It selects the features based on information gain ratio and avoids bias while selecting the features. It handles incomplete training data with missing values and has capability to use both continuous and discrete features [8]. C4.5 uses pessimistic pruning to avoid over-fitting and takes help of hill climbing algorithm to halt the process of tree generation while it meets low error rate [23].

2.3. C5.0

C5.0 is an advance version of C4.5 with additional features like boosting and unequal costs for different types of errors. It generates number of smaller trees and conducts global pruning procedure to remove sub-trees that are not helpful for improving classification accuracy [24]. It uses different weighting schemes for classifier training [9,25].

2.4. Naive Bayes tree (NBTree)

Kohavi [10] proposes a hybrid classifier called NBTree, which selects a node based on the highest utility value. It uses 5-fold cross validation value of Naive Bayes for computation of utility value at a node. NBTree is like a classical decision tree with Naive Bayes classifier at the leaf node instead of a single class.

2.5. Best-First decision tree (BFTree)

Haijian Shi [11] suggests Best-First decision tree learner, which expands best node first. That is, the node is selected based on maximum reduction of impurity among all available nodes instead of depth first order.

3. Related work

In this section, we present an overview of related vertical partitioning approaches. A partitioning method for multi-view ensemble learning [26] for both low and high dimensional scenarios is proposed. In their work, authors use homogeneous ensemble method, where same classification algorithm is used for training [26]. Kusik [27] describes partitioning techniques based on features and objects to make effective decision with respect to quality in semiconductor industry. Feature partitioning is used for classification of web pages and a Co-Training technique is used for learning with labelled and unlabelled data, The method divides the input space into independent and redundant views. Each view builds a separate model to classify unlabelled data and retain the new-data for further classification [28]. In another direction, Lior Rokach et al. [29] propose a meta classifier using feature set partitioning. Meta classifier decides whether the dataset is to be partitioned or not based on the dataset characteristics. It uses a meta-dataset to aid in partitioning a given dataset based on its learnt experience. The partitioning methods are used in mechanical design and identified partitions based on type of features (e.g nominal or ordinal) [30]. Lior Rokach et al. [14] use a feature set decomposition method, Breadth-Oblivious-Wrapper, to improve the quality in manufacturing. A theme based partitioning [15] exploits themes present in the dataset for partitioning. For instance, the themes in a Teacher dataset could be: work experience, research, skills, qualifications. Some researchers use genetic algorithm to partition the feature set and examine the effectiveness using Vapnik-Chervonenkis dimension bound [16]. Kumar and Minz [12] propose a vertical partitioning method based on Information gain and show improvement in classification accuracy of Decision tree classifier.

4. The proposed partitioning methods to decision tree classifier

In this section, we propose vertical partitioning methods, where feature set is partitioned into non-empty subsets based on the ideas of Ferrer Diagram [17] and Bell Triangle [18]. The Ferrer Diagram and Bell triangle enable us to divide the feature set non-sequentially into subsets of features with different characteristics.

Let $D = \{D_1, D_2, \ldots, D_m\}$ denotes the dataset with m features under consideration, and $C = \{C_1, C_2, \ldots, C_s\}$ denote the set of s class labels.

4.1. Decision tree using Ferrer Diagram based partitioning (DTFDP)

In this section, we propose a novel decision tree method based on the concept of Ferrer Diagram [17]. The method is illustrated in Figs. 1–2. The algorithm is given as follows:

Algorithm:

Let F = $\{F_1, F_2, \ldots, F_n\}$ be the set of features for the data under consideration, and C_i^j be the correlation coefficient of features F_i and F_j.

1. For the given training data, compute correlation coefficient, C_i^j, between the features, F_i and F_j, $\forall i, j = 1, 2, \ldots, n; i < j$ (Step-1 of Fig. 1) as given by:

$$C_i^j = \frac{cov(F_i, F_j)}{\sigma(F_i)\sigma(F_j)} \tag{1}$$

2. Find the mean of correlation values, A^j, related to a feature, F_j, $\forall j = 1, 2, \ldots, n$ (Step-2 of Fig. 1) as given by:

$$A^j = \frac{1}{n}\sum_{i=1}^{n} C_i^j \tag{2}$$

3. Arrange the features F_1 to F_n based on the ascending order of mean correlation values of features, A^1 to A^n. In other words, F_i is the feature whose mean correlation value, A^i, is the i^{th} smallest (Step-3 of Fig. 1).
4. Create M blocks, B = $\{B_1, B_2, B_3, \ldots, B_M\}$ based on the idea of Ferrer Diagram [17,19] as shown in Fig. 3 (Step-4 of Fig. 1).
5. Build a Decision Tree, R_i, for each of the M blocks, B_i, $i = 1, 2, \ldots, M$ and create an ensemble of decision trees using training data (Step-5 of Fig. 1).

$$R_i \leftarrow BuildDecisionTree(B_i) \tag{3}$$

6. A Test data instance, T, is classified using the ensemble of decision trees based on majority vote fusion (Fig. 2):
   6.1 Create M blocks $TB_1, TB_2, \ldots, TB_M$ for each test data instance, T using Step 4.
   6.2 Classify each test data block using R_i:

$$\theta_i \leftarrow R_i(TB_i) \tag{4}$$

   where θ_i stores a classification label, C_j.
   6.3 $\theta \leftarrow$ majority-vote $(\theta_1, \theta_2, \ldots, \theta_M)$, where θ stores the class label assigned to T.

Partitioning of the features (Step 4) is carried out based on the idea of Ferrer Diagram [17,19]. We illustrate this idea of partitioning for 24 and 16 features in Fig. 3. To create k partitions (blocks), it is required to arrange n features in a $\frac{n}{k} \times k$ matrix. Next, divide first $(k+1)$ rows of the matrix into R-shaped parts, where r^{th} part contains maximum of 2^r features. If the number of rows in the matrix is more than $(k+1)$ (that is, $\frac{n}{k} > (k+1)$), then repeat the process for next $(k+1)$ rows of the matrix as next pass until all the features are included in a ⌋-shaped part. We create 3 blocks for feature set $F = \{F_1, F_2, F_3, \ldots, F_{24}\}$ as $B_1 = \{F_1, F_6, F_{13}, F_{18}\}$, $B_2 = \{F_2, F_5, F_7, F_8, F_{14}, F_{17}, F_{19}, F_{20}\}$, $B_3 = \{F_3, F_4, F_9, F_{10}, F_{11}, F_{12}, F_{15}, F_{16}, F_{21}, F_{22}, F_{23}, F_{24}\}$ (Fig. 3a). The blocks obtained for another case is shown in Fig. 3b. The blocks contain features with variety of correlation values.

4.2. Decision tree using Bell Triangle based partitioning (DTBTP)

In DTFDP method, the partitioning of the feature set is performed based on the Ferrer Diagram (Step 4), whereas in DTBTP, the partitioning is carried out based on the idea of the

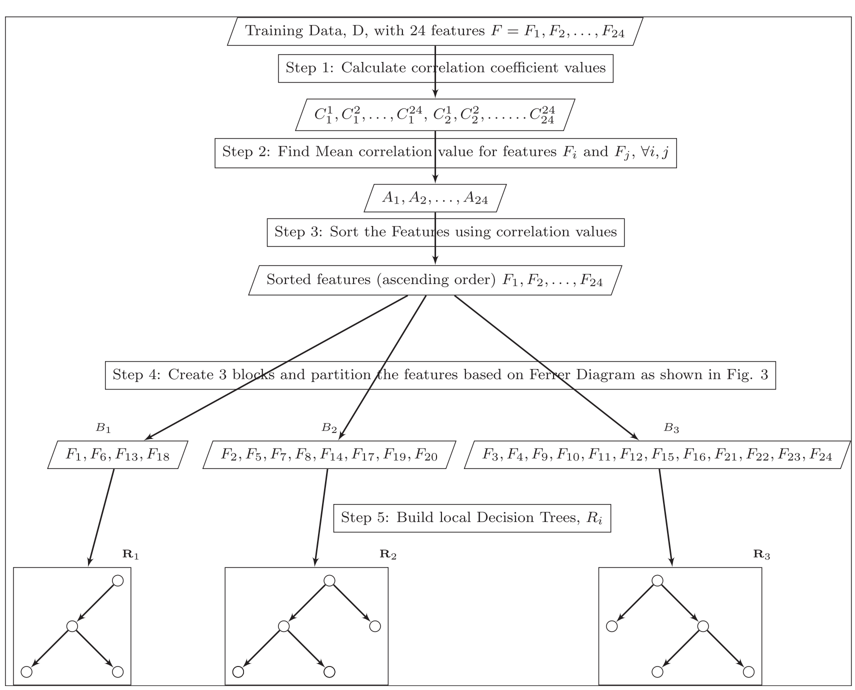


Fig. 1. Visualizing the proposed DTFDP method (Training phase).

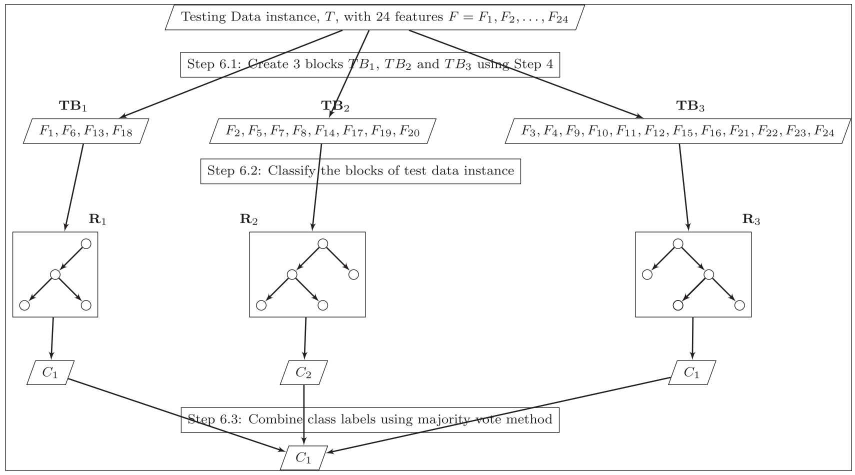
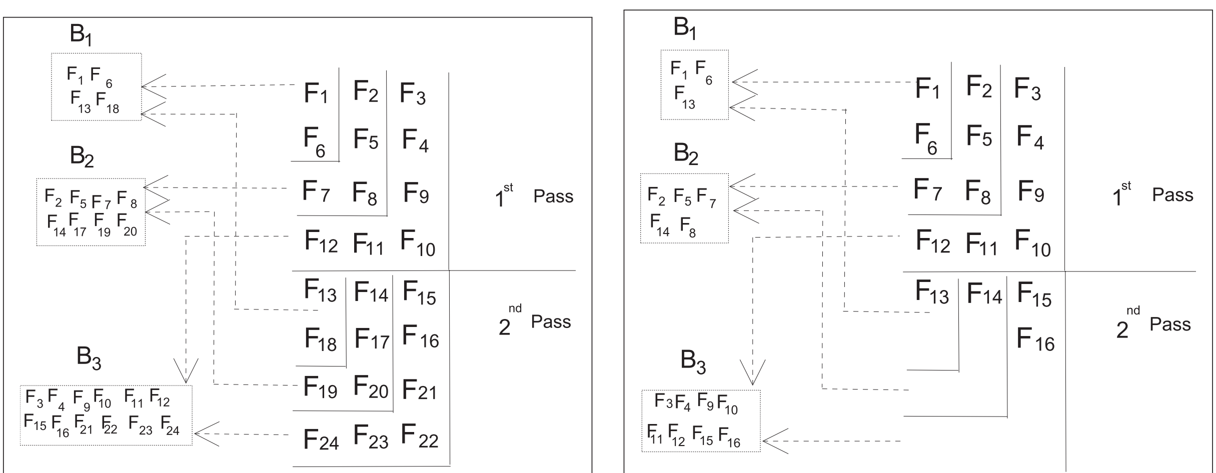


Fig. 2. Visualizing the proposed DTFDP method (Testing phase).



(a) Feature set partitioning with 24 features

(b) Feature set partitioning with 16 features

Fig. 3. Feature set partitioning based on Ferrer Diagram to create 3 blocks.

Bell Triangle [18]. Other steps are similar to DTFDP method. Every block includes features with low, medium and high correlation values. The partitioning idea of DTBTP method is illustrated for 24 and 16 features in Fig. 4. The features are arranged in the shape of a Bell triangle, that is, the r^{th} row contains exactly r features, except the last row. Next, each column of the Bell triangle is treated as a block. Single feature of the last column is merged with its immediate previous column. In this approach, the number of blocks is constant, which is equal to $c - 1$, where c is number of columns in Bell triangle.

It is to be noted that the blocks obtained by DTFDP and DTBTP methods are different.

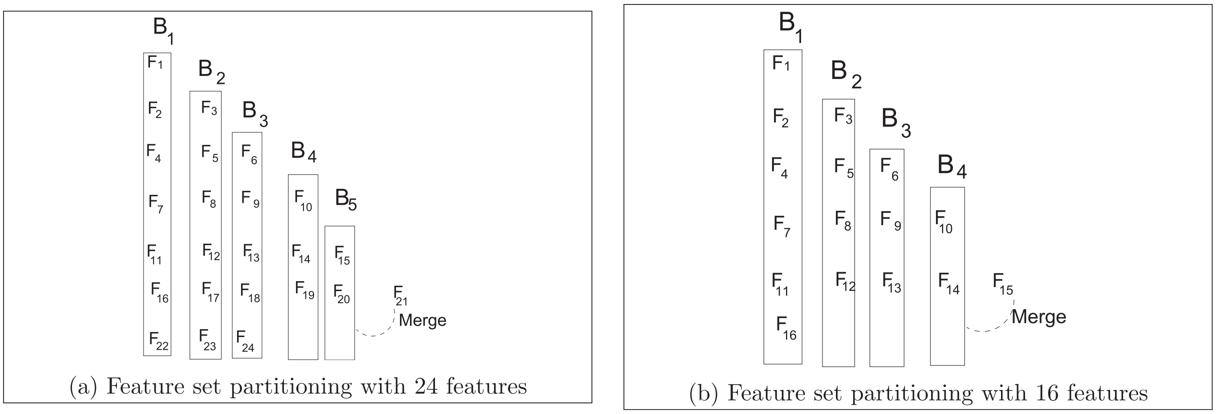
5. Results and analysis

In order to test the performance of the proposed methods, we use 11 different datasets from UCI [31] and KEEL repository [32] and compare their classification accuracy with traditional decision trees (CART, C4.5, and C5.0), NBTree, BFTree and Serial-CMFP [13]. Each dataset is vertically partitioned into 3, 4 and 5 blocks and we use 10-fold cross validation to determine the classification accuracy. Table 1 describes the number of data instances, the number of features and the number of classes for each dataset.

We use a Computer system with operating system of Windows 7, Intel i5 core processor, 8GB RAM, and R tool (version 3.1.2) [33] to compute experimental results in this work. We also use WEKA Tool [34] to perform experiments related to NBTree and BFTree.

5.1. Comparison of classification rates of DTFDP method against classical decision tree methods, NBTree, BFTree and Serial-CMFP

The classification accuracies of DTFDP + CART are computed with varied number of partitions and given in Table 2. Here, the DTFDP + CART makes use of CART method to



(a) Feature set partitioning with 24 features

(b) Feature set partitioning with 16 features

Fig. 4. Feature set partitioning based on Bell Triangle.

Table 1
Characteristics of the data sets.

Sl. No.	Data set	No. of features	No. of instances	No. of classes
1	Breast cancer Wiscosin [31]	11	699	2
2	Image segmentation [31]	20	210	7
3	Movement library [31]	91	360	15
4	Online news popularity [31]	60	39644	2
5	Parkinson disease [31]	23	195	2
6	Sonar [32]	61	208	2
7	Spambase [31]	58	4601	2
8	Thyroid [32]	22	7200	3
9	Vehicle3 [32]	19	846	2
10	Waveform [31]	22	5000	3
11	Wine [31]	14	178	3

build local Decision trees in Step 5 of section 4.1. It is observed that average classification accuracy of the DTFDP + CART method is higher than classical CART (upto 3% higher), NBTree (upto 1% higher), BFTree (upto 2% higher) and Serial-CMFP (upto 3% higher) methods (Table 2). The classification accuracy is calculated using 10 classification values obtained in 10-fold cross validation.

Table 3 summarizes the classification accuracies of DTFDP + C4.5, C4.5, NBTree, BFTree and Serial-CMFP methods. Here, the DTFDP + C4.5 method makes use of C4.5 method to build local Decision trees in Step 5 of section 4.1. It is seen that average classification accuracy of the DTFDP + C4.5 method is higher than classical C4.5 (upto 3% higher), NBTree (upto 3% higher), BFTree (upto 4% higher) and Serial-CMFP (upto 5% higher) methods (Table 3).

Table 4 presents the classification accuracies of DTFDP + C5.0, C5.0, NBTree, BFTree and Serial-CMFP methods. Here, the DTFDP + C5.0 method makes use of C5.0 method to build local Decision trees in Step 5 of section 4.1. It is seen that average classification accuracy of the DTFDP + C5.0 method is better than C5.0 (upto 2% higher), NBTree (upto 3% higher), BFTree (upto 4% higher) and Vertical-CMFP (upto 5% higher) methods (Table 4).

In brief, the proposed DTFDP method outperforms other classical methods (CART, C4.5, C5.0), NBTree, BFTree and partitioning based Serial-CMFP with partitions of varied length.

5.2. Comparison of classification rates of DTBTP method against classical decision tree methods, NBTree, BFTree and Serial-CMFP

The classification accuracies of DTBTP + CART are computed with varied number of partitions $(3, 4, 5)$ and plotted in Fig. 5. The average classification accuracy of the DTBTP + CART method is higher than classical CART (upto

Table 2
Classification accuracies of DTFDP + CART method, CART, NBTree, BFTree and Serial-CMFP.

Sl. No.	Datasets	Classification rate (%)						
		DTFDP with 3 partitions	DTFDP with 4 partitions	DTFDP with 5 partitions	CART	NBTree	BFTree	Serial-CMFP
1	Breast cancer Wiscosin [31]	96.71	95.99	96.13	94.71	96.16	95.71	97.61
2	Image segmentation [31]	88.1	89.05	86.19	84.29	80.95	84.12	76.19
3	Movement library [31]	64.72	66.67	64.17	58.33	69.44	61.11	78.7
4	Online news popularity [31]	62.65	62.24	63.07	62.8	65.61	63.48	56.24
5	Parkinson disease [31]	86.24	87.68	92.84	86.15	86.2	89.65	84.48
6	Sonar [32]	73.55	79.81	78.88	70.69	64.51	69.35	82.25
7	Spambase [31]	89.26	89.35	88.52	89.24	93.04	91.52	89.63
8	Thyroid [32]	94.53	92.58	92.9	99.57	99.3	99.58	92.77
9	Vehicle3 [32]	77.2	77.9	78.37	75.29	79.13	77.95	74.4
10	Waveform [31]	77.56	77.5	79.94	73.28	81	75.26	66.4
11	Wine [31]	94.41	91.6	93.3	82.57	88.67	84.9	83.01
Average		**82.26**	**82.76**	**83.11**	**79.72**	**82.18**	**81.14**	**80.15**

Table 3
Classification accuracies of DTFDP + C4.5 method with C4.5, NBTree, BFTree and Serial-CMFP.

Sl. No.	Datasets	Classification rate (%)						
		DTFDP with 3 partitions	DTFDP with 4 partitions	DTFDP with 5 partitions	C4.5	NBTree	BFTree	Serial-CMFP
1	Breast cancer Wiscosin [31]	95.99	95.97	96.42	93.99	96.16	95.71	97.61
2	Image segmentation [31]	87.14	88.57	86.19	87.14	80.95	84.12	76.19
3	Movement library [31]	71.39	78.06	76.94	69.72	69.44	61.11	78.7
4	Online news popularity [31]	62	63.46	64.1	58.98	65.61	63.48	56.24
5	Parkinson disease [31]	88.21	89.76	89.82	86.15	86.2	89.65	84.48
6	Sonar [32]	78.83	81.74	76.83	74.02	64.51	69.35	82.25
7	Spambase [31]	92.09	92.7	90.91	92.76	93.04	91.52	89.63
8	Thyroid [32]	94.5	92.58	93.5	99.72	99.3	99.58	92.77
9	Vehicle3 [32]	76.72	79.44	79.32	74.82	79.13	77.95	74.4
11	Waveform [31]	79.8	79.86	81.48	76.52	81	75.26	66.4
12	Wine [31]	94.41	96.63	94.97	88.78	88.67	84.9	83.01
Average		**83.73**	**85.34**	**84.58**	**82.05**	**82.18**	**81.14**	**80.15**

Table 4
Classification accuracies of DTFDP + C5.0 method with C5.0, NBTree, BFTree and Serial-CMFP.

Sl. No.	Datasets	Classification rate (%)						
		DTFDP with 3 partitions	DTFDP with 4 partitions	DTFDP with 5 partitions	C5.0	NBTree	BFTree	Serial-CMFP
1	Breast cancer Wiscosin [31]	95.7	96.13	96.71	93.99	96.16	95.71	97.61
2	Image segmentation [31]	88.57	89.05	87.14	87.14	80.95	84.12	76.19
3	Movement library [31]	73.33	78.33	77.5	74.17	69.44	61.11	78.7
4	Online news popularity [31]	65.41	65.09	65.61	64.76	65.61	63.48	56.24
5	Parkinson disease [31]	86.76	89.29	91.37	84.62	86.2	89.65	84.48
6	Sonar [32]	77.88	82.21	76.9	75.95	64.51	69.35	82.25
7	Spambase [31]	91.85	92.5	91.31	93.05	93.04	91.52	89.63
8	Thyroid [32]	94.47	92.58	93.28	99.74	99.3	99.58	92.77
9	Vehicle3 [32]	77.43	79.32	78.49	73.99	79.13	77.95	74.4
10	Waveform [31]	80.52	80.3	81.82	77.52	81	75.26	66.4
11	Wine [31]	93.89	94.38	96.11	88.21	88.67	84.9	83.01
Average		**84.16**	**85.38**	**85.11**	**83.01**	**82.18**	**81.14**	**80.15**

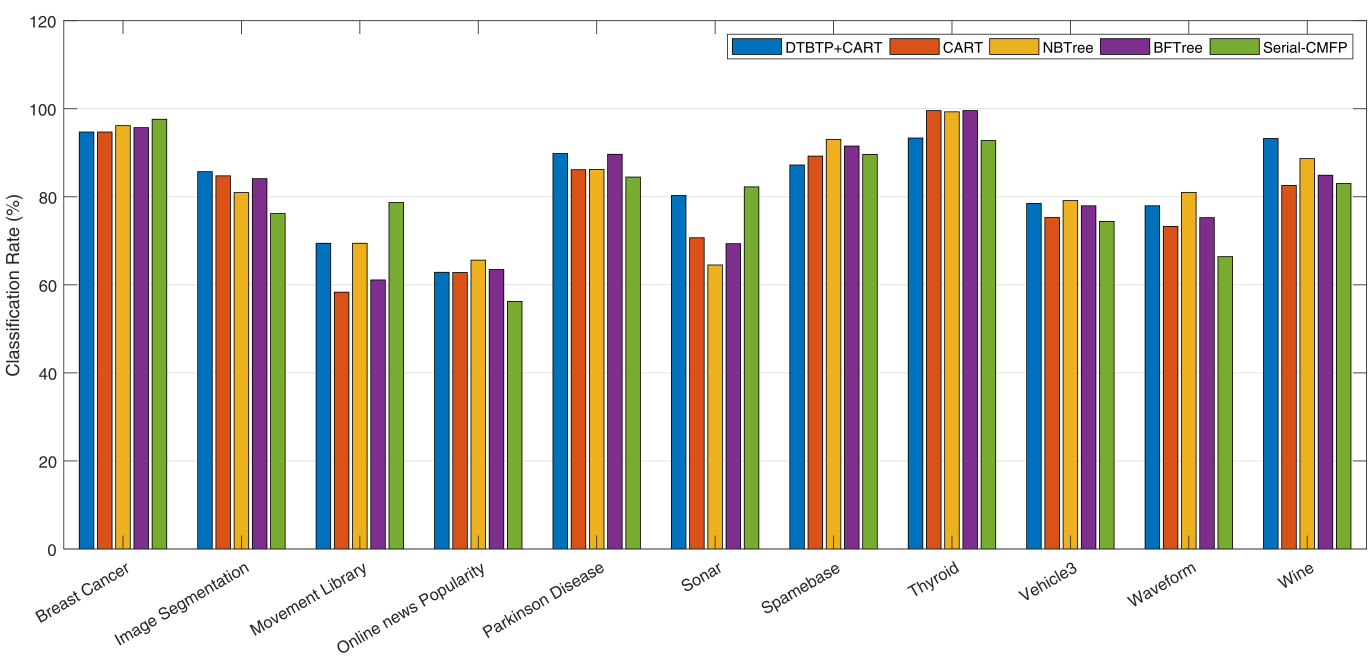


Fig. 5. Classification accuracies: DTBTP + CART method versus other methods (CART, NBTree, BFTree and Serial-CMFP).

3% higher), NBTree (upto 1% higher), BFTree (upto 2% higher) and Serial-CMFP (upto 3% higher) (Fig. 6). The average classification accuracy is computed using 10 classification values obtained in 10-fold cross validation.

Figure 7 summarizes the classification accuracies of DTBTP + C4.5, C4.5, NBTree, BFTree and Serial-CMFP methods. It is evident that average classification accuracy of the DTBTP + C4.5 method is higher than classical C4.5 (upto 3.5% higher), NBTree (upto 3% higher), BFTree (upto 4% higher) and Serial-CMFP (upto 5% higher) methods (Fig. 8).

The classification accuracies of DTBTP + C5.0, C5.0, NBTree, BFTree and Serial-CMFP methods are plotted in Fig. 9. The proposed method shows improved classification accuracy in most of the datasets. From our analysis, we observed that the average classification accuracy of the DTBTP + C5.0 method is better than C5.0 (upto 2% higher), NBTree (upto 2.5% higher), BFTree (upto 3.5% higher) and Serial-CMFP (upto 4.5% higher) methods (Fig. 10).

In a nutshell, the proposed DTBTP method shows better or competitive classification rate as compared to other classical methods for most of the datasets used.

6. Experiments related to stability

Decision tree learning algorithms are simple and easily interpreted by the users. However, they are known to be unstable as a small variation in training data instances may result in a drastic change in the tree structure and yields a different prediction. The stability of a decision tree is measured by metrics like standard deviation, dissimilarity measure, depth of the tree, number of terminal nodes and misclassification rate. There are two types of stabilities of a decision tree - Semantic

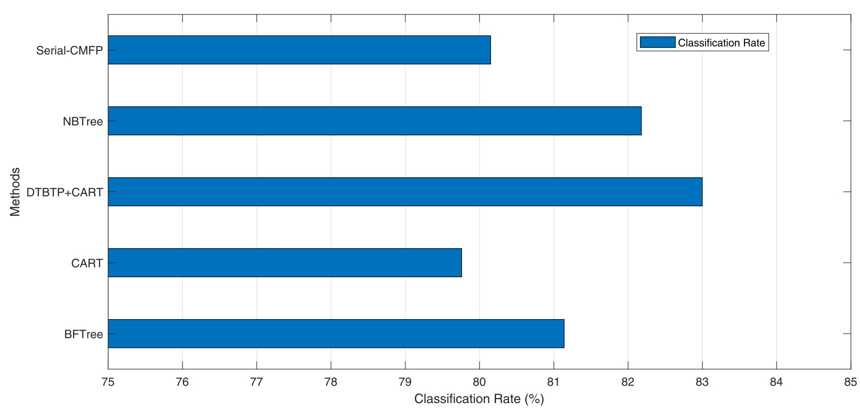


Fig. 6. Average classification accuracies: DTBTP + CART method versus other methods (CART, NBTree, BFTree and Serial-CMFP).

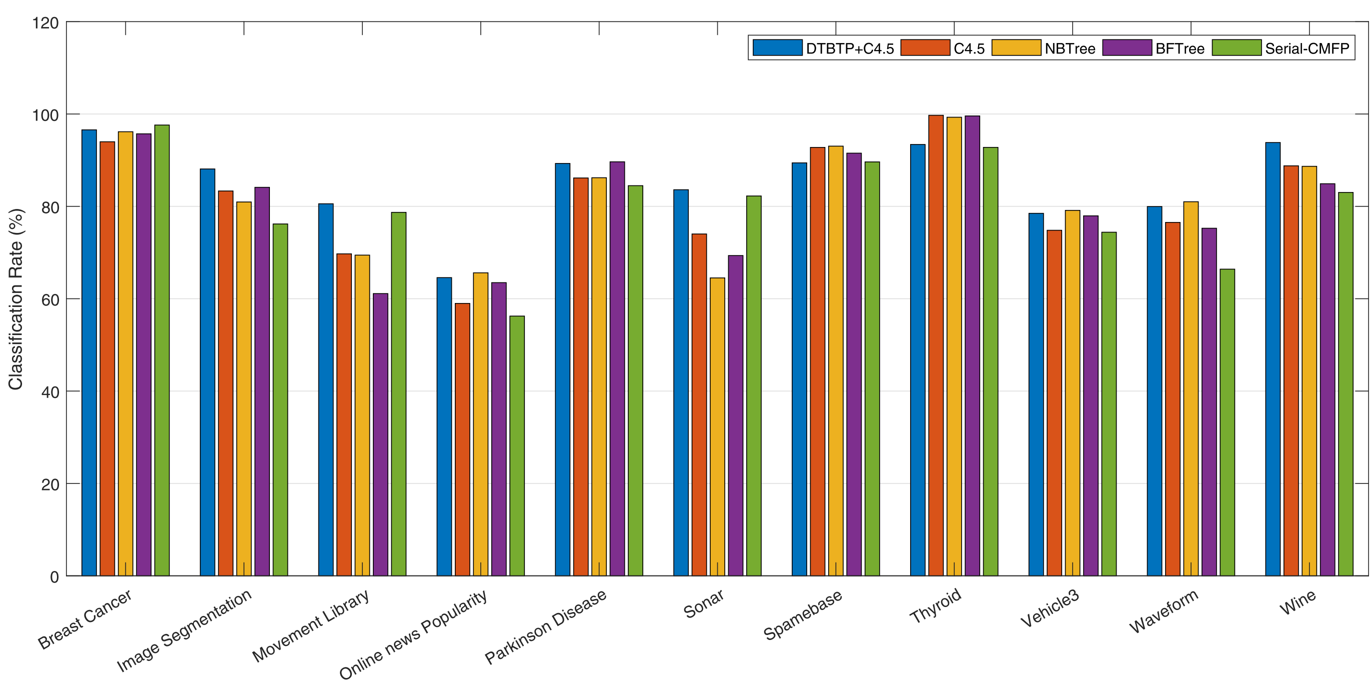


Fig. 7. Classification accuracies: DTBTP + C4.5 method versus other methods (C4.5, NBTree, BFTree and Serial-CMFP).

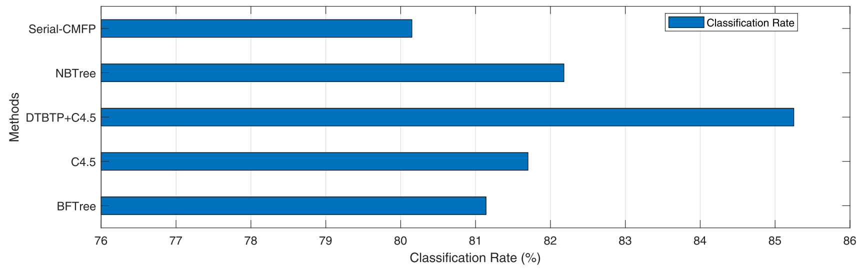


Fig. 8. Average classification accuracies: DTBTP + C4.5 method versus other methods (C4.5, NBTree, BFTree and Serial-CMFP).

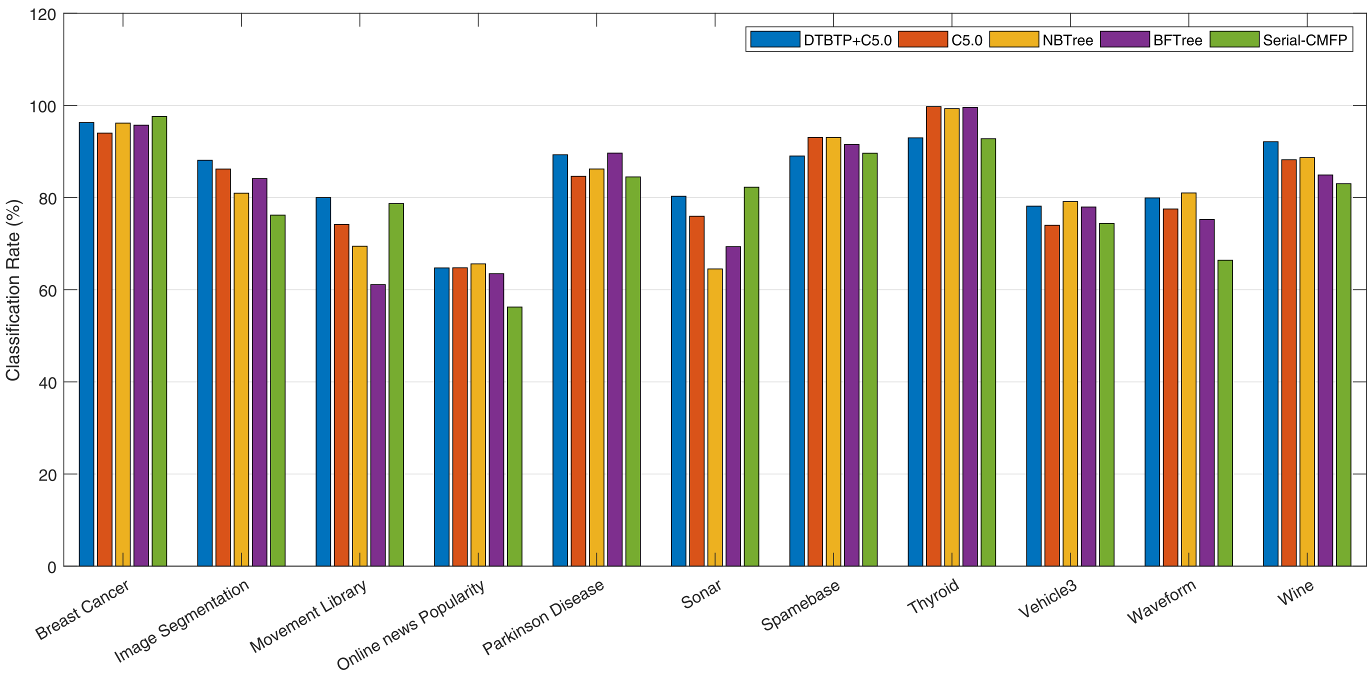


Fig. 9. Classification accuracies: DTBTP + C5.0 method versus other methods (C5.0, NBTree, BFTree and Serial-CMFP).

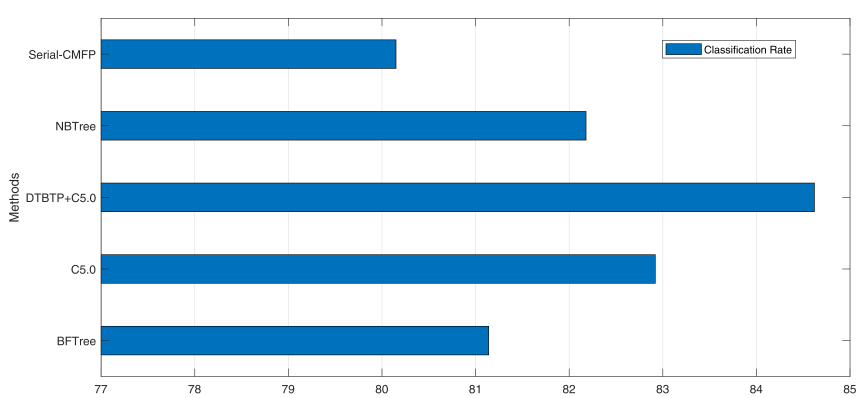


Fig. 10. Average classification accuracies: DTBTP + C5.0 method versus other methods (C5.0, NBTree, BFTree and Serial-CMFP).

and structural stabilities. Semantic stability is based on two classifiers with same prediction and structural stability is based on two classifiers with same prediction, and also have a similar topology [35].

In this section, we discuss the stability of the proposed decision tree methods using standard deviation, and misclassification rate. A method is said to be more stable, if it shows lower values of standard deviation, and misclassification rates.

6.1. Standard deviation

The standard deviation of classification rates of the proposed and other methods are tabulated in Tables 5–7.

Table 5 describes standard deviation values related to the proposed DTFDP + CART method and other methods. The proposed method shows upto 1.14% lower value over other methods. It shows lower values for 8 data sets, and higher values for Thyroid and Online news popularity datasets as compared to CART method. Table 6 shows standard deviation values of the proposed DTFDP + C4.5 and other methods. The proposed DTFDP + C4.5 shows lower standard deviation (by 1.08%) as compared to BFTree and is competitive to other methods. Table 7 describes standard deviation values of the proposed DTFDP + C5.0 and other decision tree methods. It shows an average standard deviation value of 3.57%, but traditional C5.0 yields slightly lesser value of 3.40%. It shows lower value for seven datasets as compared to traditional C5.0. A more stable method shows lower standard deviation value in classification rates. Because the proposed methods show relatively lower standard deviation values as compared to other decision tree methods, they are relatively more stable.

Table 5
Standard Deviation values of DTFDP + CART and other decision tree methods using 10 fold Cross Validation Technique.

Sl. No	Datasets	DTFDP with 5 partitions	CART	NBTree	BFTree	Serial-CMFP
1	Breast cancer Wiscosin [31]	2.16	2.63	0.72	2.47	2.79
2	Image segmentation [31]	4.73	4.91	6.52	7.09	4.01
3	Movement library [31]	5.77	7.17	7.61	8.08	4.30
4	Online news popularity [31]	0.65	0.29	2.51	0.91	0.62
5	Parkinson disease [31]	4.20	12.14	6.18	8.5	4.93
6	Sonar [32]	8.07	7.17	5.60	8.80	7.05
7	Spambase [31]	0.84	1.29	1.95	1.41	1.47
8	Thyroid [32]	0.84	0.36	1.09	0.21	1.10
9	Vehicle3 [32]	3.94	4.23	3.16	3.58	4.81
10	Waveform [31]	2.03	2.16	1.80	1.67	1.69
11	Wine [31]	5.09	8.53	4.09	7.04	7.12
Average		**3.48**	**4.62**	**3.74**	**4.59**	**3.59**

Table 6
Standard Deviation values of DTFDP + C4.5 and other decision tree methods using 10 fold Cross Validation Technique.

Sl. No	Datasets	DTFDP with 4 partitions	C4.5	NBTree	BFTree	Serial-CMFP
1	Breast cancer Wiscosin [31]	2.34	2.79	0.72	2.47	2.79
2	Image segmentation [31]	6.02	6.44	6.52	7.09	4.01
3	Movement library [31]	5.14	5.77	7.61	8.08	4.30
4	Online news popularity [31]	0.76	0.50	2.51	0.91	0.62
5	Parkinson disease [31]	5.95	4.41	6.18	8.5	4.93
6	Sonar [32]	5.91	5.65	5.60	8.80	7.05
7	Spambase [31]	1.21	1.00	1.95	1.41	1.47
8	Thyroid [32]	0.89	0.24	1.09	0.21	1.10
9	Vehicle3 [32]	3.85	5.17	3.16	3.58	4.81
10	Waveform [31]	2.68	1.09	1.80	1.67	1.69
11	Wine [31]	3.90	6.12	4.09	7.04	7.12
Average		**3.51**	**3.56**	**3.74**	**4.59**	**3.59**

Table 7
Standard Deviation values of the DTFDP + C5.0 and other decision tree methods using 10 fold Cross Validation Technique.

Sl. No	Datasets	DTFDP with 5 partitions	C5.0	NBTree	BFTree	Serial-CMFP
1	Breast cancer Wiscosin [31]	1.80	2.40	0.72	2.47	2.79
2	Image segmentation [31]	4.91	5.24	6.52	7.09	4.01
3	Movement library [31]	5.62	3.47	7.61	8.08	4.30
4	Online news popularity [31]	0.58	0.78	2.51	0.91	0.62
5	Parkinson disease [31]	5.81	4.29	6.18	8.5	4.93
6	Sonar [32]	7.46	9.64	5.60	8.80	7.05
7	Spambase [31]	1.04	1.07	1.95	1.41	1.47
8	Thyroid [32]	0.90	0.23	1.09	0.21	1.10
9	Vehicle3 [32]	4.07	4.40	3.16	3.58	4.81
10	Waveform [31]	1.81	2.12	1.80	1.67	1.69
11	Wine [31]	5.27	3.75	4.09	7.04	7.12
Average		**3.57**	**3.40**	**3.74**	**4.59**	**3.59**

6.2. Misclassification rate

Average Misclassification rates that are obtained by the various Decision tree methods for 11 datasets are shown in Figs. 11–13. From these figures, it is evident that the proposed DTFDP method exhibits lower average misclassification rate in comparison to all the other methods. Among all the methods, Serial-CMFP partitioning method and CART are most unstable as they show higher misclassification rate. It is to be noted that, the more stable classifier has lower misclassification rate. Therefore, the proposed method shows improved stability based on the misclassification rate.

6.3. Why do proposed methods show improved classification rate over other methods (CART/C4.5/C5.0, NBTree, BFTree and Serial-CMFP)?

The proposed methods show better classification results, as compared to other methods (CART/C4.5/C5.0, NBTree, BFTree and Vertical-CMFP) because of the following rea-

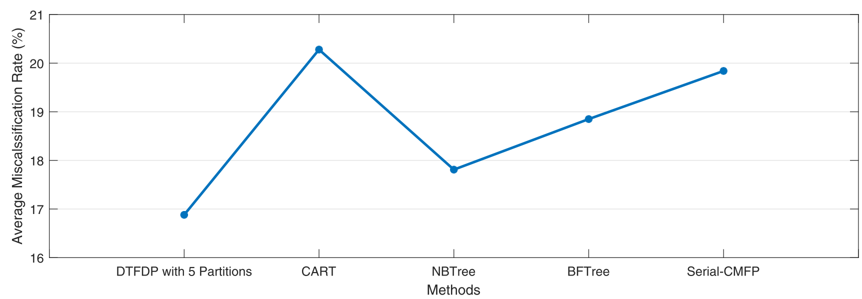


Fig. 11. Average misclassification rates of DTFDP + CART and other decision trees (CART, NBTree, BFTree and Serial-CMFP).

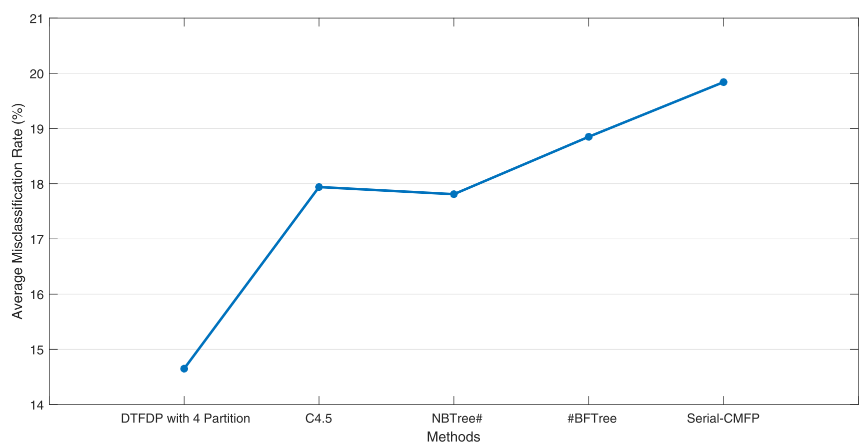


Fig. 12. Average misclassification rates of DTFDP + C4.5 and other decision trees (C4.5, NBTree, BFTree and Serial-CMFP).

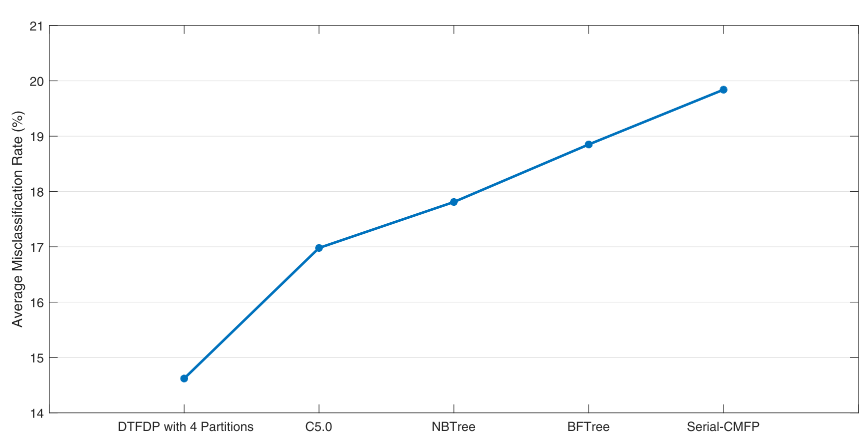


Fig. 13. Average misclassification rates of DTFDP + C5.0 and other decision trees (C5.0, NBTree, BFTree and Serial-CMFP).

sons: (i) The proposed methods make use of subset of features for construction of decision trees. Because the size of feature subset is far lower than the number of data instances as compared to the size of feature set, the curse of dimensionality is reduced; (ii) The proposed methods make use of features with variety of characteristics (like low to high correlation values), which may lead to building more accurate local decision trees; (iii) The proposed methods are relatively more robust to outliers, because the effect of outlier features is confined to a block; (iv) The proposed methods make use of non-sequential ways of creating feature blocks to create a better mixture of features with variety of correlation values.

7. Conclusion and future work

In this paper, we proposed Decision tree methods, using non-sequential ways of partitioning, that show significant improvement in terms of classification accuracy over classical decision tree techniques, NBTree, BFTree and Serial-CMFP. Each partition includes variety of features with low, medium and high correlation values. The proposed methods have been proved to be more stable as compared to other methods using the measures of standard deviation and misclassification rates.

Acknowledgements

The authors are grateful to Technical Education Quality Improvement Programme (TEQIP-III) of B.M.S. College of Engineering, Bengaluru for providing sufficient facilities to publish this work.

References

[1] Lee Jae Sung, Lee Eun Sung. Exploring the usefulness of a decision tree in predicting peoples location. In: Procedia − social and behavioural sciences; 2014. p. 447−51.

[2] Natek Srecko, Zwilling Moti. Student data mining solution knowledge management system related to higher education institutions. Comput Ind 2014;41:6400−7.

[3] Jayasree Vikas, Siva Balan RV. Money laundering regulatory risk evaluation using Bitmap Index-based Decision Tree. J Assoc Arab Univ Basic Appl Sci 2017;23:96−102.

[4] Oliveira Werbeston D, Vieira Joo PA, Bezerra Ubiratan H, Martins Daniel A, Rodrigues Benedito das G. Power system security assessment for multiple contingencies using multiway decision tree. Elec Power Syst Res 2017;148:264−72.

[5] Liu Chang, Hu Zhenhua, Li Yan, Liu Shaojun. Forecasting copper prices by decision tree learning. Resour Pol 2017;52:427−34.

[6] Jing Li, Rebecca J, Stones, Gang Wang, Xiaoguang Liu, Zhongwei Li, et al. Hard drive failure prediction using Decision Trees. Reliab Eng Syst Saf 2017;164:55−65.

[7] Breiman L. Classification and regression trees. Wadsworth Int. Group; 1984.

[8] Dai Wei, Ji Wei. A MapReduce implementation of C4.5 decision tree algorithm. Int J Database Theor Appl 2014;7:49−60.

[9] C5.0, See5: an informal tutorial. 1993. http://www.rulequest/see5-win.html.

[10] Kohavi Ron. Scaling up the accuracy of naive-bayes classifiers: a decision-tree hybrid. In: Proceedings of the 2nd international conference on knowledge discovery and data mining. Portland; 1996. p. 202−7.

[11] Shi H. Best-first decision tree learning. Master's thesis. Hamilton, NZ: University of Waikato; 2007.

[12] Kumar Vipin, Minz Sonajharia. Multi-view ensemble learning: a supervised feature set partitioning for high dimensional data classification. In: Proceedings of the Third International Symposium on women in computing and informatics; 2015. p. 31−7.

[13] Seetha Hari, Narasimha Murty M. Classification by majority voting in feature partitions. Int J Inf Decis Sci 2016;8:109−24.

[14] Rokach Lior, Maimon Oded. Data mining for improving the quality of manufacturing: a feature set decomposition approach. J Intell Manuf 2006;17:285−99.

[15] Kadappa Vijayakumar, Guggari Shankru, Negi Atul. Decision tree classifier using theme based partitioning. In: IEEE International Conference on computing and network communications (CoCoNet'15); 2015. p. 546−52.

[16] Rokach Lior. Genetic algorithm-based feature set partitioning for classification problems. J Pattern Recogn Soc 2008;41:1676−700.

[17] Comtet L. Advanced combinatorics: the art of finite and infinite expansions, rev. enl. ed.. D.Reidel Publishing Company; 1973.

[18] Todd Tichenor. Bounds on graph compositions and the connection to the Bell triangle. Discrete Math 2016;339:1419−23.

[19] Franceschini Fiorenzo, Maisano Domenico. The citation triad: an overview of a scientists publication output based on Ferrers diagrams. J Inf 2010;4:503−11.

[20] Anderson Thomas R, hankin Robin KS, Killworth Peter D. Beyond the Durfee square: enhancing the h-index to score total publication output. Scientometrics 2008;3:577−88.

[21] Yang Winston. Bell numbers and k-trees. Discrete Math 1996;156:247−52.

[22] Rokach L. Decision tree. In: Book chapter, Vol. 24; July 2001.

[23] Wu Xindong, Kumar Vipin, Ross Quinlan J, Ghosh Joydeep, Yang Qiang, Motoda Hiroshi, et al. Top 10 algorithms in data mining. Knowl Inf Syst 2008;14:1−3.

[24] Kuhn M, Johnson K. Applied predictive modeling. Springer; 2013.

[25] Efe Akka, Akin Lutfiye, Evren ubuku H, Artuner Harun. Application of Decision Tree Algorithm for classification and identification of natural minerals using SEMEDS. Comput Geosci 2015;80:38−48.

[26] Kumar Vipin, Minz Sonajharia. Multi-view ensemble learning using optimal feature set partitioning: an extended experiments and analysis in low dimensional scenario. In: Second International Symposium on computer vision and the internet (VisionNet15)Vol. 58; 2015. p. 499−506.

[27] Kusiak A. Decomposition in data mining: an industrial case study. In: IEEE trans electron packag manuf; 2000. p. 345−53.

[28] Blum A, Mitchell T. Combining labeled and unlabeled data with cotraining. In: COLT: proceedings of the workshop on computational learning theory. Morgan Kaufmann Publishers; 1998.

[29] Rokach Lior. Decomposition methodology for classification tasks − a meta decomposer framework. Pattern Anal Appl 2006;9:257−71.

[30] Kusiak A, Larson N. Decomposition and representation methods in mechanical design. J Mech Des 1995;117(B):17−24.

[31] Aha D, Murphy P. UCI Repository of machine learning databases. Irvine, CA: University of California, Department of Information and Computer Science; 1994. http://www.ics.uci.edu/mlearn/MLRepository.html.

[32] Alcal-Fdez J, Fernandez A, Luengo J, Garca S, Snchez L, Herrera F. KEEL data-mining software tool: data set repository, integration of algorithms and experimental analysis framework. J Mult Valued Log Soft Comput 2011;17:255−87.

[33] R, the R project for statistical computing. 1993. http://www.r-project.org/.

[34] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The weka data mining software: an update. ACM SIGKDD Explor Newsl 2009;11(1):10−8.

[35] Mirzamomen Z, Kangavari MR. A framework to induce more stable decision trees for pattern classification. Pattern Anal Appl 2016:1−14.