Advances in Science, Technology and Engineering Systems Journal Vol. 2, No. 3, 845-854 (2017)

www.astesj.com

Special Issue on Recent Advances in Engineering Systems

ASTESJ

ISSN: 2415-6698

Performance Evaluation of Associative Classifiers in Perspective of Discretization Methods

Zulfiqar Ali[1,2*], Waseem Shahzad[1]

[1]*Department of Computer Science, National University of Computer & Emerging Science, Islamabad, ZIP Code #44000, Pakistan.*

[2]*Department of Computer Science and IT, The University of Lahore, Lahore, ZIP Code#54000, Pakistan.*

A R T I C L E I N F O

Article history:
Received: 22 April, 2017
Accepted: 16 May, 2017
Online: 19 June, 2017

Keywords:
Associative Classification
Discretization Methods
KEEL
Data Mining

A B S T R A C T

Discretization is the process of converting numerical values into categorical values. Contemporary literature study reveals that there are many techniques available for numerical data discretization. The performance of classification method is dependent on the exploitation of the data discretizing method. In this article, we investigate the effect of discretization methods on the performance of associative classifiers. Most of the classification approaches work on the discretized databases. There are various approaches exploited for the discretization of the database to compare the performance of the classifiers. The selection of the discretization method greatly influences the classification performance of the classification method. We compare the performance of associative classifiers namely CBA and CBA2 on the selective discretizing methods i.e. 1R Discretizer (1R-D), Ameva Discretizer (Ameva-D), Bayesian Discretizer (Bayesian-D), Discretization algorithm based on Class-Attribute Contingency Coefficient (CACC-D), Class-Attribute Dependent Discretizer (CADD-D), Distribution-Index-Based Discretizer (DIBD-D), Cluster Analysis (ClusterAnalysis-D), Chi-Merge Discretizer (ChiMerge-D) and Chi2 Discretizer (Chi2-D) in terms of accuracy. The main object of this study is to investigate the impact of discretizing method on the performance of the Associative Classifier by keeping constant other experimental parameters. Our experimental results show that the performance of the Associative Classifier significantly varies with the change of data discretization method. So the accuracy rate of the classifier is highly dependent on the selection of the discretization method. For this comparative performance study, we use the implementation of these methods in KEEL data mining tool on public datasets.

1. Introduction

Discretization methods have played a great role in data mining and knowledge discovery. The discretization process makes learning more accurate and faster. There are various emerging classification problems in various domains of knowledge like image processing, medical science, business analytics and data mining etc. data, images, audio, video and textual data. The rapid growth in the data reservoirs in the fields of business, basket analysis, Engineering sciences, social networks, stock exchange and geological data is very high due to the cheaper storage resources. The high growth rate and huge data volume create a challenging problem i.e. knowledge discovery from the huge databases in the field of Data Mining. For the appropriate, effective and comprehensive knowledge discovery for the managers and decision makers; researchers are proposing continuously more efficient knowledge mining approaches.

The field of artificial neural networks, expert systems, medical science, bioinformatics, machine learning is example areas where extensively classification approaches have been studied. There are various approaches exploited for the building of associative classifiers. This comparative study provides the extension of the work presented in [1]. We provide the performance analysis of Associative Classifiers (CBA and CBA2) with the variation of the data discretizing method.

Thabtahand Fadi Abdeljaber provided the review of associative classification in [2]. Ranjana Vyas et al. describe the application of Associative Classifiers for Predictive analytics in [3]. The exploitation of associative classifiers for predictive analysis in the field of health care is surveyed in [4] by Sunita Soni and O.P.Vyas. Huan Liu et al. provide the extensive survey of the

*Zulfiqar Ali, National University of Computer & Emerging Science, H-11/4, Islamabad, Pakistan, +923335262172 & zulfiqar.ali@nu.edu.pk

discretization techniques in[5]. The focus of this article was an exploration of discretization methods with prospective of their historic development, the trade-off between speed and accuracy. Authors provided the hierarchical framework to categorize the exiting discretizing methods. Sotiris Kotsiantis and Dimitris Kanellopoulos surveyed the discretization techniques applied for the data discretization in [6].The theoretical and empirical perspective of the discretizing methods is very in [7] by the Salvador Garcıa. In [8] Pancho et al. provided the analysis of fuzzy association rules with Fingrams in KEEL.

In this article, we investigate the effect of discretizing methods on the performance of Associative Classifiers CBA and CBA2. We use the selective data discretizing methods i.e. 1R Discretizer (1R-D) [9], Ameva Discretizer (Ameva-D) [10], Bayesian Discretize r(Bayesian-D) [11], Discretization algorithm based on Class-Attribute Contingency Coefficient (CACC-D) [12], Class-Attribute Dependent Discretizer (CADD-D) [13], Distribution-Index-Based Discretizer (DIBD-D) [14], Cluster Analysis (ClusterAnalysis-D) [15], Chi-Merge Discretizer (ChiMerge-D) [16] and Chi2 Discretizer (Chi2-D) [17] for continuous data discretization purpose by exploiting the implementation of these methods in KEEL [18], a data mining tool by using the public datasets.Our experimental results reveal that the performance of the Associative Classifier significantly varies with the change of data discretization method in terms of accuracy. So the accuracy rate of the classifier is highly dependent on the selection of the data discretizing method. Our comparative study reveals that the performance of CBA (Associative Classifier) is better on the Ameva Discretizer than the other discretizing methods in terms of accuracy. The main object of this study is to investigate the impact of discretizing method on the performance of the Associative Classifier at the same other experimental parameters.

The Section 2 of the paper discusses the associative classification and describes the selective associative classification methods that are the focus of our study for the comparative analysis. Section 3 describes the data discretization process and selective methods for the data discretization i.e. 1R-D, Ameva-D, Bayesian-D, CACC-D, CADD-D, DIBD-D, ClusterAnalysis-D, ChiMerge-D, and Chi2-D. Section 4 explains the experimental Set-up exploited for this study, data sets and KEEL tool used for the experimentation. Section 5 describes the comparative performance results achieved by various discretizing methods used for datasets discretization by using Associative Classifiers CBA and CBA2. In section 6 more results discussion is provided and finally, the last Section concludes the study.

2. Associative Classification

The Associative Classification (AC) is a classification approach which integrates the classification rules mining and association rules mining that are two important data mining tasks. The Association Rule Mining (ARM) is unsupervised learning method in which no class attribute involved during the discovery of rules. The aim of the association rule mining is to discover associations between items in a transaction database. The attributes in the consequent of a rule could be more than one in association rule mining. The associative classification is a supervised leaning where a class must be given for the discovery of classification rules. For the construction of a classifier that can forecast the classes of test data objects is the main objective of associative classification. The consequent of a rule is an only class attribute. The over fitting is a considerable issue in the associative classification rule discovery. The over view of the selective Associative Classification approach CBA which is exploited to investigate the impact of discretizing method on the performance of the classification approach is given in the following sections.

2.1. CBA

Bing Liu, Wynne Hsu and Yiming Ma proposed a new hybrid classification approach by integrating the concept of association rule mining and classification rule mining in [19] that is named Classification Based on Associations (CBA). In this associative classification approach, the integration is done by focusing on the discovery of a special subset of association rules that are known as class association rules (CARs).

All class association rules are discovered those satisfy the minimum support and minimum confidence by using an existing association rule mining algorithms[20].The CBA associative classifier consists of two parts 1) a rule generator (CBA-RG) and 2) a classifier builder (CBA-CB).This approach possesses various advantages like the discretization of continuous attributes based on the classification pre-determined class target. The Data Mining task in CBA consists of the three steps;1) discretization of continuous attributes if any;2) class association rules;3) classifier building based on the generated class association rules.

2.2. CBA2

Bing Liu, Yiming Ma and Ching-Kian Wong proposed the enhancement and improvements in an associative classifier CBA. The new improved associative classification approach is named CBA2 developed in [21]. In this paper, theauthor tried to coup up with weaknesses of an exhaustive search based classification system CBA. The authors proposed two new techniques to deal with the observed weaknesses of the classification approaches. The first weakness observed is that as the traditional association rule mining exploits only a single minsup in rule generation which results inadequate for unbalanced class distribution. Secondly, classification data often contains a huge number of rules, which may cause a combinatorial explosion. For various databases, the rule generator is unable to generate rules with many conditions while such rules may be important for accurate classification. The first problem with this approach is tackled by using multiple class minsups in rule generation instead of single minsup as in CBA. The second problem which is caused by the exponential growth of the number of rules is dealt indirectly. The decision tree method [22] is exploited. The main working concept of the CBA2 is to use the rules of CBA2 to segment the training data and then select the classifier. These improvements in CBA improved the accuracy and lower error rate of the classification.

3. Data Discretization

Discretization is a data preprocessing technique used in many knowledge discoveries, machine learning and data mining tasks. Discretization process converts the continuous data into discrete form as most of the knowledge discovery and data mining algorithms work on discrete data. The discretization technique transforms a set of continuous attributes into discrete ones. By associating categorical values to intervals discretizing approach transforms quantitative data into qualitative data. The data discretiztion techniques are exploited to enhance the performance of the many knowledge discoveries and data mining approaches. We have used selective discretization method for our study to investigate the performance of Associative Classifier CBA by using public data sets. The discretization methods used in this study are described in the following.

3.1. Discretization Methods

Gonzalez-Abrilet al. proposed a new discretization method named Ameva Discretizer (Ameva-D) in [10] that is proposed for supervised learning algorithms. The Ameva discretization approach maximizes a contingency coefficient based on Chi-square statistics. It helps in generating a potentially minimal number of discrete intervals. The most distinguishing feature of Ameva with respect to other discretizing approaches is that it does not need the user to indicate the number of intervals.

Xidong Wu proposed a new discretizing algorithm namely Bayesian Discretizer (Bayesian-D) in [11]. The Bayesian-D discretization approach exploits the Bayes formula. Cheng-Jung Tsai et al. propose a discretization algorithm based on Class-Attribute Contingency Coefficient named (CACC-D) in [12]. The CACC-D discretizing approach is motivated by the contingency coefficient. The CACC algorithm is a static, global and incremental discretizing approach. The CACC-D is supervised and top-down discretization algorithm which is based on Class-Attribute Contingency Coefficient. J.Y. Ching et.al proposed anew method for continuous data discretization named Class-Attribute Dependent Discretizer (CADD-D) in [13]. The class-dependant discretizing is performed in this approach for inductive learning from continuous and mix-mode data. The CADD-D is a discretizing method optimized for supervised learning based on information theoretic discretization method. The interdependence redundancy between the discrete intervals and the class labels is measured in the CADD-D method tom maximize the mutual dependence.

L.A. Kurgan et.al proposed a discretization method namely known as Class-Attribute Interdependence Maximization (CAIM-D) in [23]. The CAIM-D method is proposed for the supervised data classification. The main objective of this proposed approach is to increase the class-attribute interdependence to maximum level and to produce a minimal number of discrete intervals. The classification results in terms of accuracy are more promising with CAIM discretization method with respect to other discretization approaches. Huan Liu and Rudy Setiono proposed a method for converting numeric data discrete named Chi2 Discretizer (Chi2-D) in [17]. This approach (Chi2-D) takes data sets with numeric attributes as an input. This approach can intelligently and automatically discretize the numeric attributes as well as remove irrelevant ones as output. The Chi2 algorithm applies the X 2 statistic which conducts a significance test on the relationship between the values of an attribute and the categories.

In [16] Randy Kerber proposed a method for the discretization known as Chi-Merge Discretizer (ChiMerge-D) which is a general, robust algorithm that uses the X 2 statistic to discretize numeric attributes. The ChiMerge approach provides a useful and reliable summarization of numeric attributes. The number of intervals needed is determined according to the characteristics of the data. Michal R. Chmielewski and Jerzy W. Grzymala-Busse proposed discretizing method based on Cluster Analysis named ClusterAnalysis-D in [15]. The hierarchical cluster analysis is used for discretizing attributes in ClusterAnalyusis-D.

The ClusterAnaysis-D can be classified as either locally discretizing method or globally discretizing method. The methods that are characterized by operating only one attribute are called local method while the methods considering all attributes are called global methods. QingXiang Wu et.al proposed a new discretizing approach named Distribution-Index-Based Discretizer(DIBD-D) in [14]. The DIBD-D approach is based on the definition of dichotomic entropy and a compound distribution index. This criterion is applied to discretize continuous attributes adaptively. The DIBD-D can discretize any continuous attribute adaptively according to the simple adaptive rules. The adaptive rule is based on maximal compound decrement and minimal dichotomic entropy.

4. Experimental Set-Up

In this section, we conduct experiments to evaluate the performances of the associative classification systems. For the comparative performance analysis of the selective associative classifiers, we exploited the implementations of these algorithms included in Knowledge Extraction based on Evolutionary Learning (KEEL) [18]. The overview of the Data Mining and machine learning tool KEEL is given in the following section. In this section, we describe the datasets used for the comparative analysis of the associative classifier (CBA) and CBA2 in terms of accuracy and error rate by using the different discretizing methods. The parameters set for the experiments and the experiment graph designed for these experiments in the KEEL tool are described in this section.

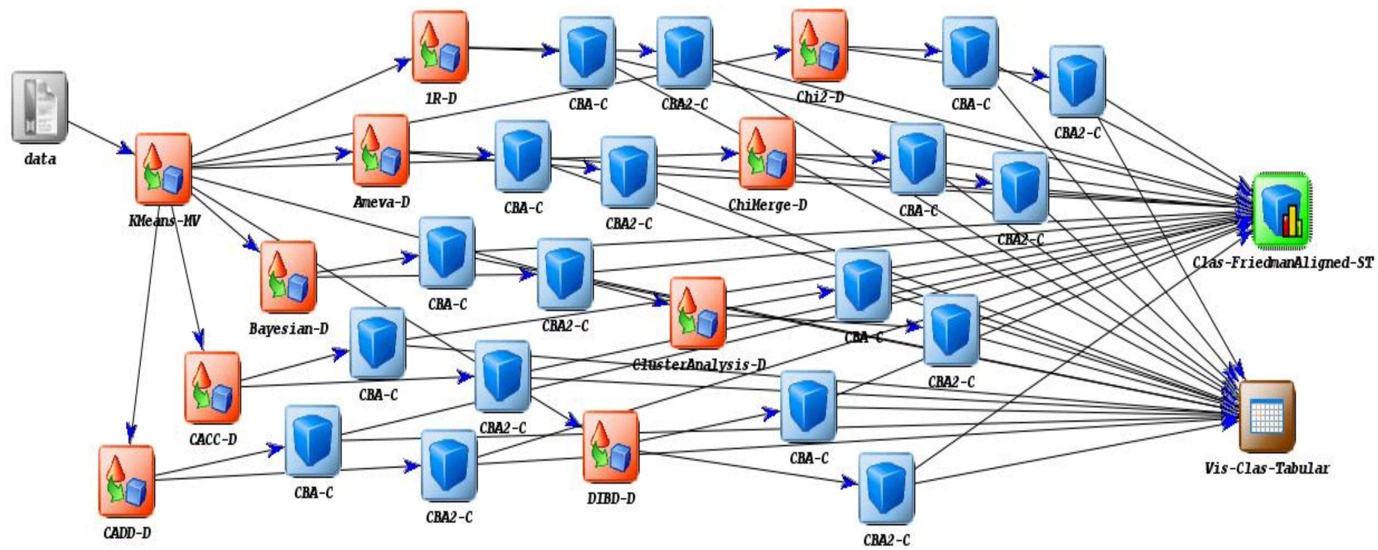
4.1. Data Sets

The description of datasets used for the comparative performance analysis of the selective Associative Classifiers under this study is given in Table2. The number of attributes (#Attributes), number of instances in the database (#Examples) and thenumber of classes (#Classes) are shown in the table. The missing values (Missing_V) in the dataset are representing by "Yes" (missing values present)/ "No" (missing values not present). The missing values of the datasets are imputed with the KMean-MV module implemented in KEEL. The datasets are discretized with the Ameva-D module included in KEEL as the associative classifiers accept the discretized form of datasets. We use the 10-fold cross-validation model for the datasets provided in KEEL. Table I summarizes the main characteristics of the 12 datasets which are given at Knowledge Extraction based on Evolutionary Learning (KEEL)-dataset repository[18].

Table 1 Data Sets Considered For The Experimental Stu

Dataset Name	#Attributes	#Examples	#Classes	Missing_V
Bupa	6	345	2	No
Cleveland	13	297	5	yes
Ecoli	7	336	8	No
Glass	9	214	7	No
Harberman	3	306	2	No
Iris	4	150	3	No
Monks	6	432	2	No
Newthyroid	5	215	3	No
Pima	8	768	2	No
Vehicle	18	846	4	No
Wine	13	178	3	No
Wisconsin	9	683	2	Yes

Figure 1 Experiment Graph Generated in KEEL



4.2. Experiment Graph

The experiment graph shows the components of the experiment and describes the relationships between them. The experimental graph of the comparative study is given in Figure1. The first component of the experimental graph is data which enables to select the datasets given in the KEEL Tool as well as to load user datasets. In our study, we selected standard KEEL datasets. The second component of the graph is KMeans-MV which is a module to impute the missing values in the database. The third component of the experiment graph is amodule for data discretization. In our case, we use the selective nine modules i.e.1R-D, Ameva-D, Bayesian-D, CACC-D, CADD-D, DIBD-D, ClusterAnalysis-D, ChiMerge-D and Chi2-D for the discretization of continuous data values. The fourth stage of the experiment graph is Associative Classification methods (CBA and CBA2) which are the focus of our study. The last stage of the experiment graph is the modules for the representation of the results of theclassifier and astatistical module for the analysis of the results produced by the algorithms used in the experiment. The module Vis-Class-Tabular provides the facility of representation of results of multiple classification methods in the form tabular representation.

4.3. Parameters of the Methods

The parameters of the associative classifiers (CBA and CBA2) under the focus of this comparative study are shown in the Tabel.2. The parameters of the method are selected according to the recommendation of the corresponding authors which are the default parameters settings included in the KEEL software tool [18]. In the Table 2,*Minsup* stands for minimum support, *Minconf* for minimum confidence, and *RuleLimit* for maximum candidate rules limit in the corresponding method.

Table 2 Parameters of the Methods for Experiment

Methods	Parameters
CBA-C	*Minsup = 0.01, Minconf = 0.5, Pruned = yes, RuleLimit = 80,000*
CBA2-C	*Minsup = 0.01, Minconf = 0.5, Pruned = yes, RuleLimit = 80,000*

5. Experimental Results

Table3 shows the comparative performance of the selected Associative Classifiers. We use the implementation of the corresponding algorithms in KEEL 3.0 for our comparative performance analysis of various discretizing methods by using CBA and CBA2 Associative Classifiers on the public datasets. We used the selective 9 discretization methods i.e 1R-D, Ameva-D, Bayesian-D, CACC-D, CADD-D, DIBD-D, ClusterAnalysis-D, ChiMerge-D and Chi2-D for this comparative study. The performance of both Associative Classifiers CBA and CBA2 is investigated on 12 public datasets for 9 discretizing methods. In table 3, the fold face red values show the best performance of CBA classifier for a specific dataset at a particular discretizing method with respect to other discretizing methods under the focus of this study. The fold face blue values in table 3 represents the best performance of the CBA2 classifier on a specific dataset with the exploitation of a specific discretizing method with respect to other discretizing methods. The last row of table 3 shows the average performance of the corresponding classifier for all datasets for a specific discretizing method. Overall the best average performance of classifier for a specific increasing method is presented with bold face red and blue for CBA and CBA2 respectively in table 3.The performance of the CBA classifier is better on Ameva-D discretizing method as compare to other discretizing methods but the average performance is better on the ChiMerge-D. The performance of CBA2 is promising with ChiMerge-D discretizing method. The performance of each discretizing method varies with respect to change in the database. The values bold face shows the wining of the corresponding discretizing methods for the CBA and CBA2 on the corresponding datasets.

Table 3 The Comparative Performance Results Of Associative Classification Methods

	1R-D		Ameva-D		Bayesian-D		CACC-D		CADD-D		DIBD-D		ClusterAnalysis-D		ChiMerge-D		Chi2-D	
Datasets	**CBA**	**CBA2**	**CBA**	**CBA2**	**CBA**	**CBA2**	**CBA**	**CBA2**	**CBA**	**CBA2**	**CBA**	**CBA2**	**CBA**	**CBA2**	**CBA**	**CBA2**	**CBA**	**CBA2**
Bupa	59.3746	59.2388	**66.9423**	**67.3425**	64.2971	60.2366	66.2854	64.4565	57.8924	57.8242	65.6888	65.6911	61.1710	62.6642	62.2117	62.2369	63.2472	65.0482
Cleveland	51.4624	45.3568	54.3978	53.7048	**58.0968**	47.1554	54.1613	54.0469	52.8172	48.6315	54.4409	53.7341	54.8065	50.7429	52.0968	**54.3304**	54.7957	53.1574
Ecoli	72.6114	67.8334	**78.0214**	72.1925	72.6292	66.9989	77.3886	**76.7704**	49.1176	48.3795	76.4973	76.4868	71.4082	66.5046	75.9002	72.1682	77.1658	74.6070
Glass	59.0015	57.9856	50.3569	49.7675	54.2565	54.8018	35.5661	35.4949	52.7213	53.1945	65.4916	60.8186	54.3821	52.6976	67.4854	**66.3038**	**70.3353**	65.6928
Haberman	74.1720	73.8807	**74.7742**	**74.1349**	71.5376	71.4663	74.1505	72.9814	73.5269	73.2942	73.4839	73.8416	73.1935	72.1017	73.4731	73.5386	72.2043	68.2111
Iris	91.3333	90.9091	92.6667	92.7273	93.3333	93.9394	92.6667	92.7273	74.0000	76.9697	91.3333	89.6970	88.0000	86.6667	92.0000	91.5152	**95.3333**	**95.7576**
Monks	52.6608	52.6255	**97.2674**	**97.5159**	51.2972	51.3858	**97.2674**	**97.5159**	52.6608	52.6255	58.5273	59.4050	80.6451	80.3385	**97.2674**	**97.5159**	**97.2674**	**97.5159**
New-Thyroid	**94.0043**	93.2900	93.9610	94.0968	92.5758	92.4242	93.0303	93.2507	87.3377	86.7965	92.1212	91.5978	93.9827	**94.5494**	93.0303	93.6836	92.1212	92.8571
Pima	73.1810	70.0638	71.2276	72.1981	**73.7160**	**73.3807**	72.3948	70.6555	65.1081	65.0924	72.4051	72.1998	69.6657	71.8346	72.7980	72.7916	72.9212	72.4235
Vehicle	67.3768	63.0685	**70.7955**	70.1350	68.4398	66.2605	68.5672	69.7059	56.9636	60.1401	64.7745	64.8752	62.5364	61.7571	67.0168	68.7395	68.4412	**70.8913**
Wine	91.4706	91.2359	94.9020	**94.8604**	93.8235	93.3749	93.7255	93.7908	80.3268	77.0648	92.0915	91.2953	**95.4575**	94.3553	93.8562	92.8699	82.5163	80.0654
Wisconsin	81.6936	82.0403	96.1366	95.9627	95.1304	95.3096	96.1366	95.9627	65.5238	65.6277	96.1387	95.7030	94.5652	94.6640	**96.4244**	**96.4860**	**96.4224**	96.3523
Average	**72.3619**	70.6274	**78.4541**	77.8865	**74.0944**	72.2278	**76.7784**	76.4466	**63.9997**	63.8034	**75.2495**	74.6121	**74.9845**	74.0730	**78.6300**	**78.5150**	**78.5643**	77.7150

Table 4 Win/Draw/Lose Record of Discretization Methods for CBA and CBA2 Classifiers

Discretizing Methods	**1R-D**		**Ameva-D**		**Bayesian-D**		**CACC-D**		**CADD-D**		**DIBD-D**		**ClusterAnalysis-D**		**ChiMerge-D**		**Chi2-D**	
Classifiers	**CBA**	**CBA2**	**CBA**	**CBA2**	**CBA**	**CBA2**	**CBA**	**CBA2**	**CBA**	**CBA2**	**CBA**	**CBA2**	**CBA**	**CBA2**	**CBA**	**CBA2**	**CBA**	**CBA2**
Win	1	0	**4**	**3**	2	1	0	1	0	0	0	0	1	1	0	**3**	2	2
Draw	0	0	**1**	**1**	0	0	1	1	0	0	0	0	0	0	2	**1**	2	1
Loss	11	12	**7**	**8**	10	11	11	10	12	12	12	12	11	11	10	**8**	8	9

Table 4 shows the comparative performance record of CBA and CBA2 in terms of Win/Draw/Lose of the Discretizing Methods. The 1R-D, Ameva-D, Bayesian-D, CACC-D, CADD-D, DIBD-D, ClusterAnalysis-D, ChiMerge-D, and Chi2-D. The performance of CBA classifier is better on Ameva-D and Chi2-D discretized. The performance of CBA2 remains same for Ameva-D and ChiMerge-D discretizing methods as shown in table 4.

6. Results Discussion

6.1. Comparative Performance Analysis in terms of Accuracy

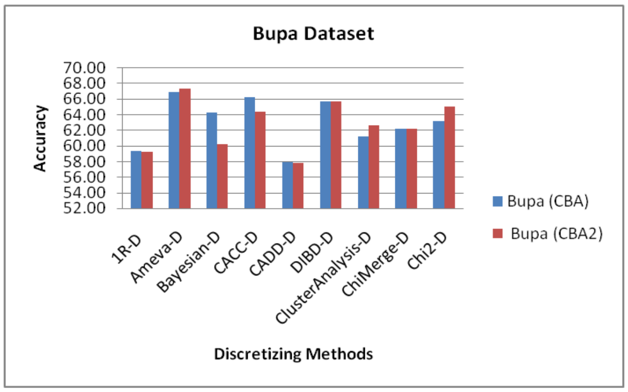
Figure2 shows the comparative performance of the CBA and CBA2 on different discretizing methods for the selective datasets. With the critical observation of the graphs in figure 2, the performance behavior of CBA and CBA2 is symmetric in terms of accuracy for various discretizing methods. The performance of Associative Classifiers (CBA and CBA2) significantly changes with the change of discretizing method as shown in figure 2. The performance variation of CBA and CBA2 at datasets Glass, Monks, Ecoli and Wisconsin at the usage of different discrete methods is given in figure2. At the dataset Glass in figure 2, there is significant variation in performance of Associative Classifiers (CBA and CBA2). The performances of CBA and CBA2 are highest for Chi-D and lowest for CACC-D discretizing methods. The experiments for the dataset Ecoli shows that the performances of CBA and CBA2 lower significantly at the usage of CADD-D discretized while remaining almost same for the other discretizing methods. For dataset Monks in figure 2, the performances of Classifiers are drastically decreasing for 1R-D, Bayesian-D and CADD-D discretizing methods. The performance of CBA2 significantly decreased on dataset Haberman for Chi2-D discretizing method. Finally, the experimental results at Wine and Wisconsin datasets, reveals that the performances of CBA and CBA2 significantly down for the CADD-D discretized in terms of accuracy.

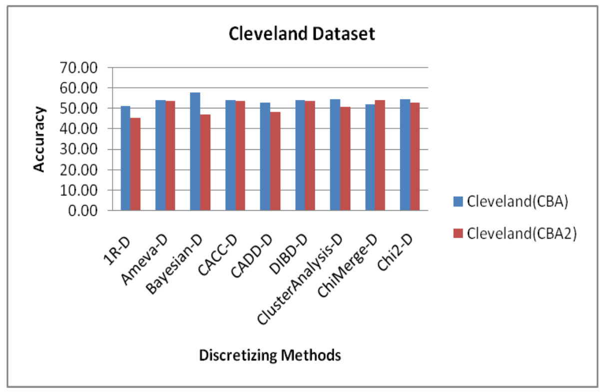
6.2. Comparative Performance Analysis in terms of Variance

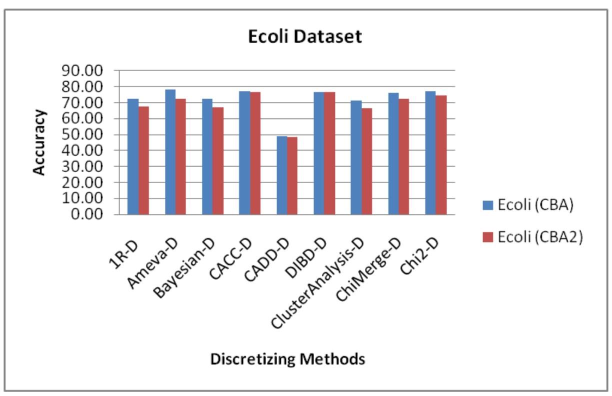
This subsection describes the impact of discretizing methods on the performance of Associative Classifiers i.e. CBA and CBA2 in terms of variance on public datasets. The variance depicts the consistency of the classification approach. The lower value of the variance indicates that the classifier is more consistent for the corresponding dataset. The discretizing methods producing a smaller value of variance for a specific dataset for the specific classification approach are more promising and provide more robustness for the classifier. The variation of performance of

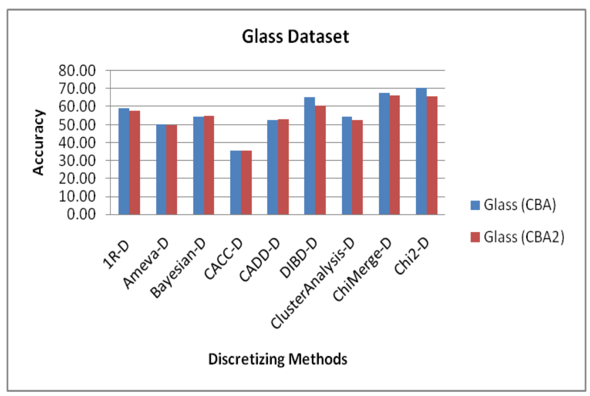
Associative Classifiers in terms of variance is elaborated in Figure 3. The performance of both classifiers is almost same for the datasets Bupa and Cleveland for on all the discretizing methods.Variance is significantly high for Chi2-D and Ameva-D discretizing methods for CBA2 and CBA classifiers on datasets Bupa and Cleveland respectively. The performance associative classifiers (CBA and CBA2) for Iris, New-thyroid, Wine and Wisconsin datasets is very promising and consisting in terms of variance for all the discretizing methods except CADD-D, Chi2 and 1R-D. The variance is very high of CBA and CBA2 on Iris and New-thyroid for CADD-D discretizing method.

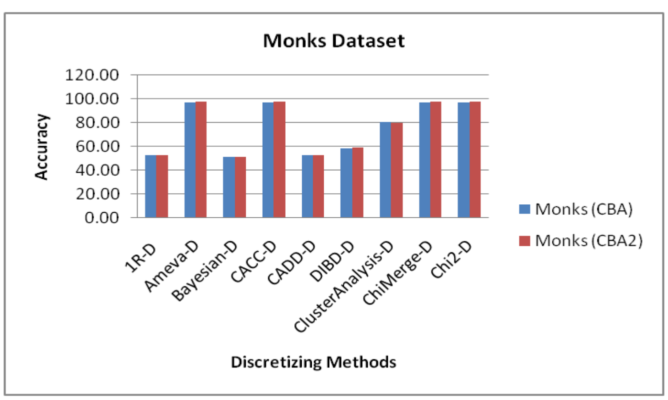
With the critical observation of Figure 3, it concluded there is a significant difference in the performance of associative classifiers in terms of variance as well as for the various discretizing methods on the pubic datasets under the focus of this study. Mostly of the discretizing methods produce promising results for some datasets while on the other datasets their performance is lower. The performance of classification approach significantly is dependent on the discetzing method used for the discretizing the continuous datasets in discrete form.

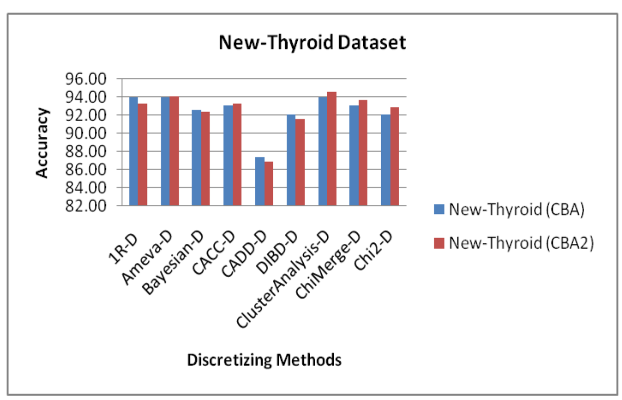












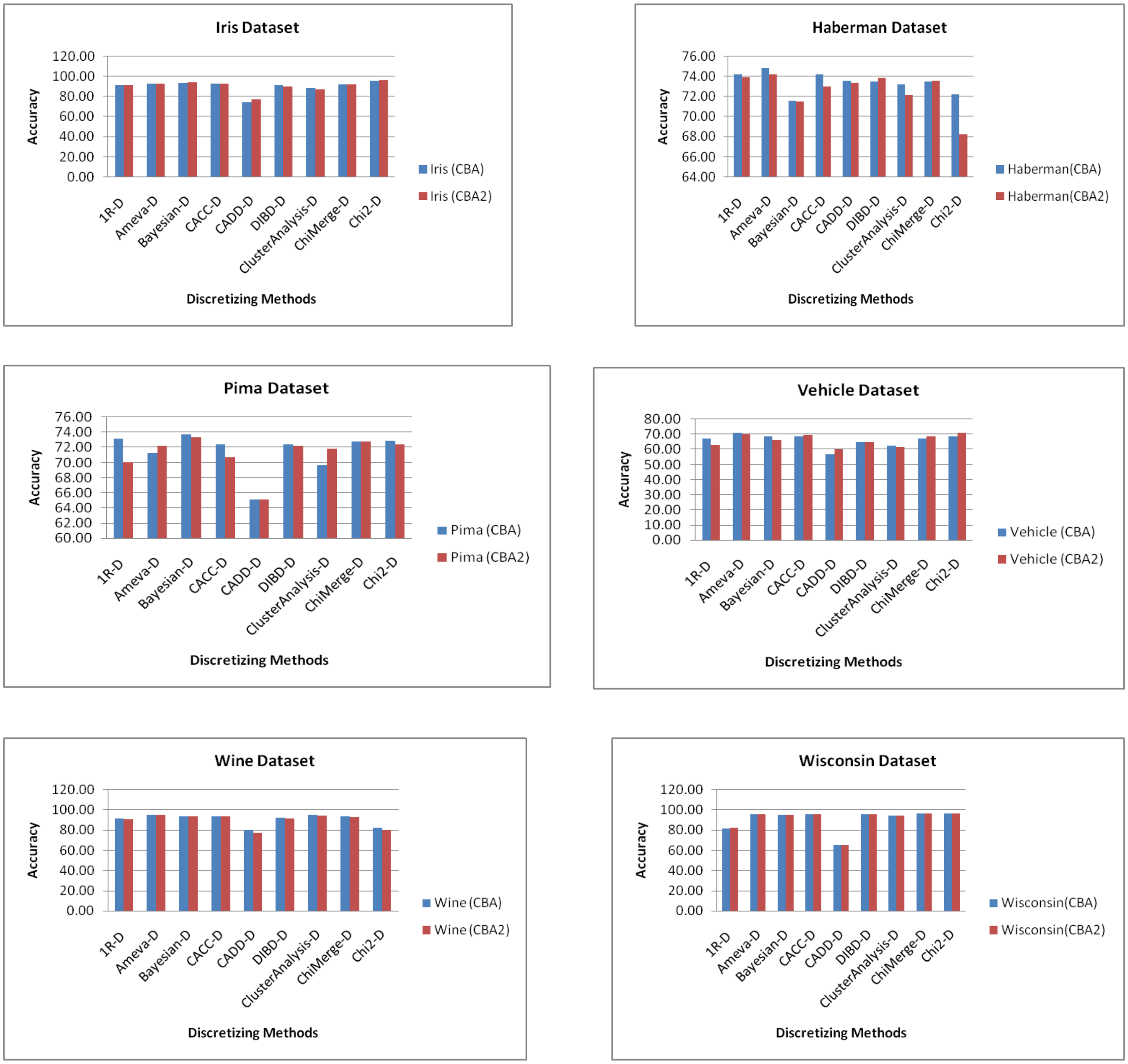
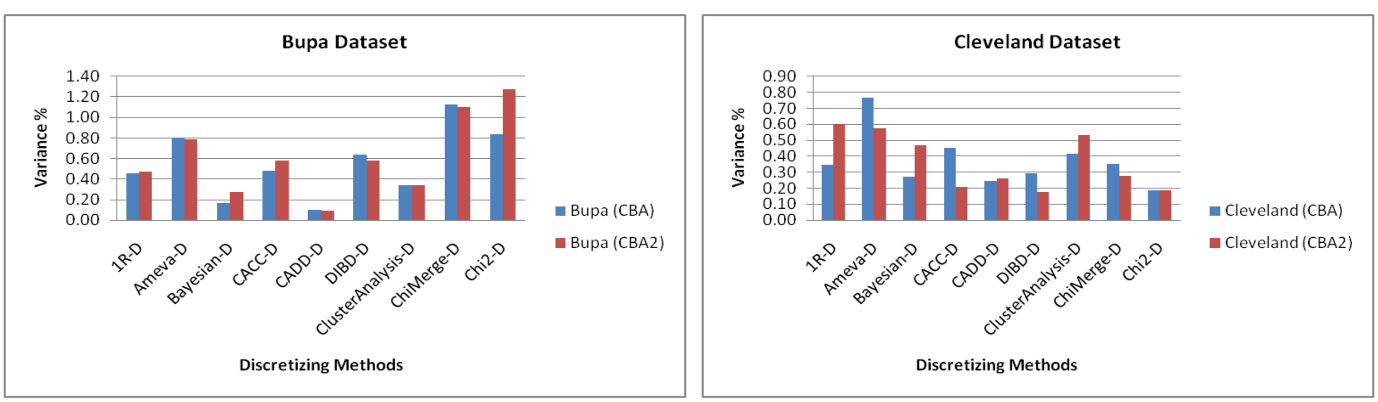
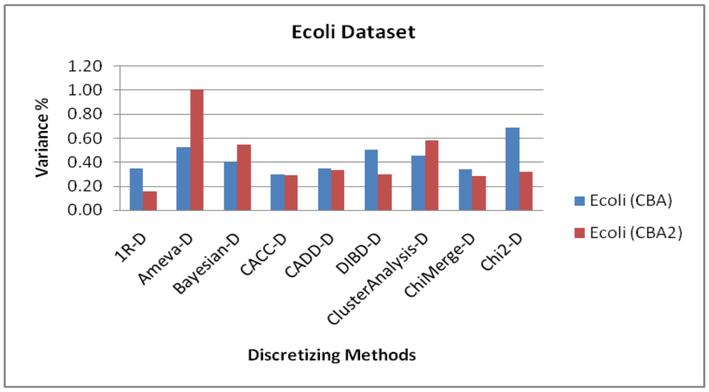
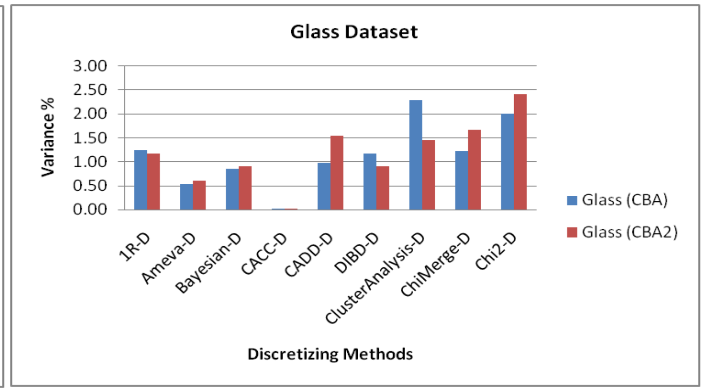
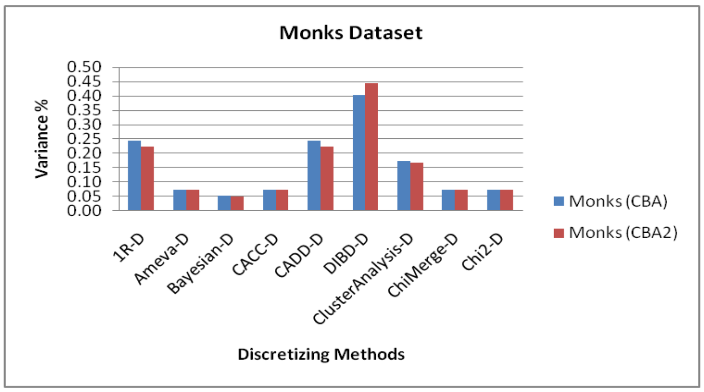


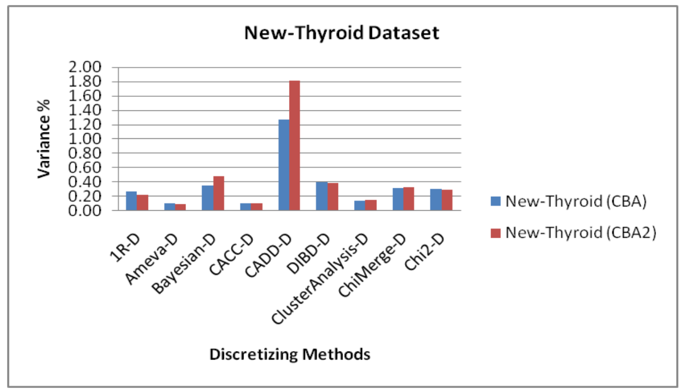
Figure 2 Impact of Discretization Methods on the Performance Associative Classifiers (CBA and CBA2) on Various Datasets

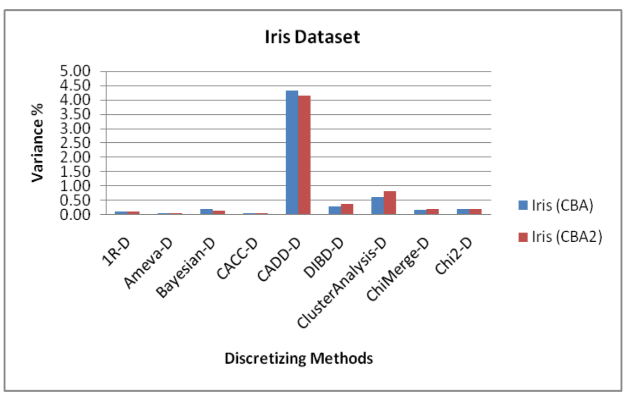


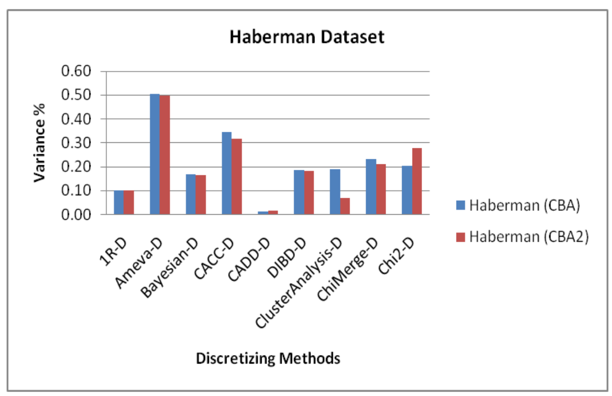

Ecoli Dataset
Variance %
1.20
1.00
0.80
0.60
0.40
0.20
0.00
1R-D
Ameva-D
Bayesian-D
CACC-D
CADD-D
DIBD-D
ClusterAnalysis-D
ChiMerge-D
Chi2-D
Ecoli (CBA)
Ecoli (CBA2)
Discretizing Methods

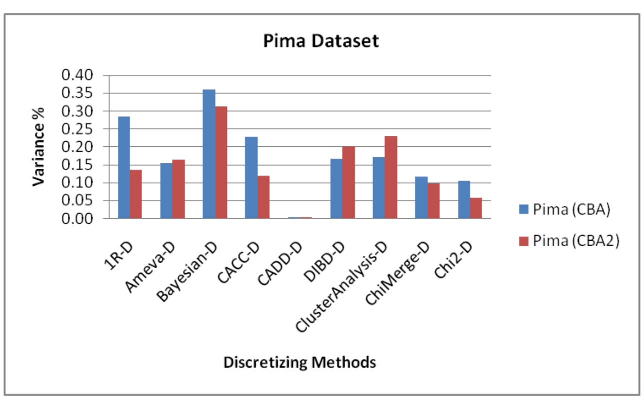

Glass Dataset
Variance %
3.00
2.50
2.00
1.50
1.00
0.50
0.00
1R-D
Ameva-D
Bayesian-D
CACC-D
CADD-D
DIBD-D
ClusterAnalysis-D
ChiMerge-D
Chi2-D
Glass (CBA)
Glass (CBA2)
Discretizing Methods

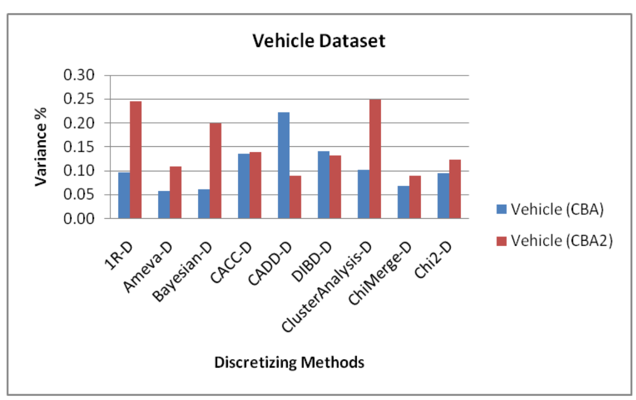

Monks Dataset
Variance %
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1R-D
Ameva-D
Bayesian-D
CACC-D
CADD-D
DIBD-D
ClusterAnalysis-D
ChiMerge-D
Chi2-D
Monks (CBA)
Monks (CBA2)
Discretizing Methods

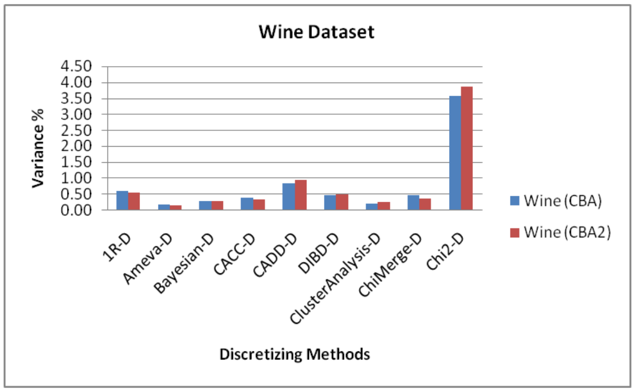

New-Thyroid Dataset
Variance %
2.00
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
1R-D
Ameva-D
Bayesian-D
CACC-D
CADD-D
DIBD-D
ClusterAnalysis-D
ChiMerge-D
Chi2-D
New-Thyroid (CBA)
New-Thyroid (CBA2)
Discretizing Methods


Iris Dataset
Variance %
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
1R-D
Ameva-D
Bayesian-D
CACC-D
CADD-D
DIBD-D
ClusterAnalysis-D
ChiMerge-D
Chi2-D
Iris (CBA)
Iris (CBA2)
Discretizing Methods


Haberman Dataset
Variance %
0.60
0.50
0.40
0.30
0.20
0.10
0.00
1R-D
Ameva-D
Bayesian-D
CACC-D
CADD-D
DIBD-D
ClusterAnalysis-D
ChiMerge-D
Chi2-D
Haberman (CBA)
Haberman (CBA2)
Discretizing Methods


Pima Dataset
Variance %
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1R-D
Ameva-D
Bayesian-D
CACC-D
CADD-D
DIBD-D
ClusterAnalysis-D
ChiMerge-D
Chi2-D
Pima (CBA)
Pima (CBA2)
Discretizing Methods


Vehicle Dataset
Variance %
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1R-D
Ameva-D
Bayesian-D
CACC-D
CADD-D
DIBD-D
ClusterAnalysis-D
ChiMerge-D
Chi2-D
Vehicle (CBA)
Vehicle (CBA2)
Discretizing Methods



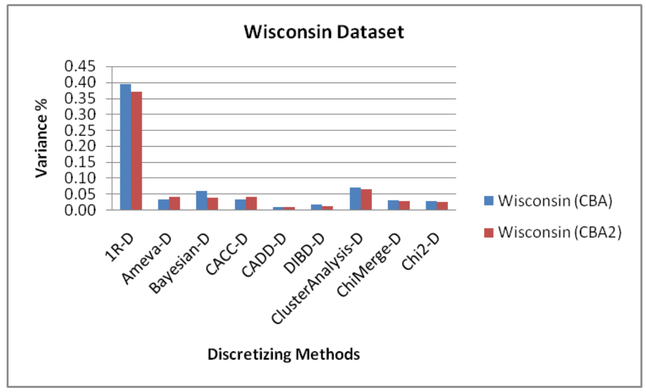


Figure 3 Comparative Impact of Discretizaing Methods on the Performance of Associative Classifiers (CBA and CBA2) in terms of Variance

7. Conclusion

Discretization algorithms have played a great role in the performance of classification techniques. We investigate the effect of discretization Methods on the Performance of Associative Classifiers. Most of the classification approaches work on the discretized databases. There are various approaches exploited for the discretization of the database to compare the performance of the classifiers. The selection of the discretization method influences the classification performance of the classification method. We compare the performance of associative classifiers namely CBA and CBA2 on the selective discretizing methods i.e. 1R-D, Ameva-D, Bayesian-D, CACC-D, CADD-D, DIBD-D, ClusterAnalysis-D, ChiMerge-D and Chi2-D in terms of accuracy and error rate. Our experimental results show that the performance of the Associative Classifiers significantly varies for different discretization methods for the same classifier. So the accuracy rate and variance in results of the classifier is highly dependent on the choice of the discretization method. For this comparative performance study, we use the implementation of these methods in KEEL data mining tool on public datasets.

In future, we will analyse the impact of discretizing methods for other classifiers by considering other parameters and to derive the significance of results by using statistical methods.

Conflict of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The authors wish to thanks Higher Education Commission of Pakistan. This work is supported in part by a grant from Higher Education Commission of Pakistan.

References

[1] Ali, Z. and W. Shahzad. *Comparative Study of Discretization Methods on the Performance of Associative Classifiers*. in *Frontiers of Information Technology (FIT), 2016 International Conference on*. 2016. IEEE.

[2] Thabtah, F., *A review of associative classification mining.* The Knowledge Engineering Review, 2007. **22**(01): p. 37-65.

[3] Vyas, R., et al. *Associative classifiers for predictive analytics: Comparative performance study*. in *Computer Modeling and Simulation, 2008. EMS'08. Second UKSIM European Symposium on*. 2008. IEEE.

[4] Soni, S. and O. Vyas, *Using associative classifiers for predictive analysis in health care data mining.* International Journal of Computer Applications, 2010. **4**(5): p. 33-37.

[5] Liu, H., et al., *Discretization: An enabling technique.* Data mining and knowledge discovery, 2002. **6**(4): p. 393-423.

[6] Kotsiantis, S. and D. Kanellopoulos, *Discretization techniques: A recent survey.* GESTS International Transactions on Computer Science and Engineering, 2006. **32**(1): p. 47-58.

[7] Garcia, S., et al., *A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning.* IEEE Transactions on Knowledge and Data Engineering, 2013. **25**(4): p. 734-750.

[8] Pancho, D.P., et al. *Analyzing fuzzy association rules with Fingrams in KEEL.* in *2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. 2014. IEEE.

[9] Holte, R.C., *Very simple classification rules perform well on most commonly used datasets.* Machine learning, 1993. **11**(1): p. 63-90.

[10] Gonzalez-Abril, L., et al., *Ameva: An autonomous discretization algorithm.* Expert Systems with Applications, 2009. **36**(3): p. 5327-5332.

[11] Wu, X., *A Bayesian discretizer for real-valued attributes.* The Computer Journal, 1996. **39**(8): p. 688-691.

[12] Tsai, C.-J., C.-I. Lee, and W.-P. Yang, *A discretization algorithm based on class-attribute contingency coefficient.* Information Sciences, 2008. **178**(3): p. 714-731.

[13] Ching, J.Y., A.K.C. Wong, and K.C.C. Chan, *Class-dependent discretization for inductive learning from continuous and mixed-mode data.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995. **17**(7): p. 641-651.

[14] Wu, Q., et al., *A distribution-index-based discretizer for decision-making with symbolic ai approaches.* IEEE transactions on knowledge and data engineering, 2007. **19**(1): p. 17-28.

[15] Chmielewski, M.R. and J.W. Grzymala-Busse, *Global discretization of continuous attributes as preprocessing for machine learning.* International journal of approximate reasoning, 1996. **15**(4): p. 319-331.

[16] Kerber, R. *Chimerge: Discretization of numeric attributes*. in *Proceedings of the tenth national conference on Artificial intelligence*. 1992. Aaai Press.

[17] Liu, H. and R. Setiono, *Feature selection via discretization.* IEEE Transactions on knowledge and Data Engineering, 1997. **9**(4): p. 642-645.

[18] Alcalá, J., et al., *Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework.* Journal of Multiple-Valued Logic and Soft Computing, 2010. **17**(2-3): p. 255-287.

[19] Ma, B.L.W.H.Y. and B. Liu. *Integrating classification and association rule mining*. in *Proceedings of the 4th*. 1998.

[20] Agrawal, R. and R. Srikant. *Fast algorithms for mining association rules*. in *Proc. 20th int. conf. very large data bases, VLDB*. 1994.

[21] Liu, B., Y. Ma, and C.-K. Wong, *Classification using association rules: weaknesses and enhancements.* Data mining for scientific applications, 2001. **591**.

[22] Salzberg, S.L., *C4. 5: Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993.* Machine Learning, 1994. **16**(3): p. 235-240.

[23] Kurgan, L.A. and K.J. Cios, *CAIM discretization algorithm.* IEEE transactions on Knowledge and Data Engineering, 2004. **16**(2): p. 145-153.