Avrupa Bilim ve Teknoloji Dergisi
Sayı 35, S. 275-281, Nisan 2022
© Telif hakkı EJOSAT'a aittir
Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
No. 35, pp. 275-281, April 2022
Copyright © 2022 EJOSAT
Research Article

Comparison of Discretization Methods for Classifier Decision Trees and Decision Rules on Medical Data Sets

Yılmaz Kaya[1], Ramazan Tekin[2*]

[1] Siirt University, Faculty of Engineering, Departmant of Computer Engineering, Batman, Turkey, (ORCID: 0000-0001-5167-1101), yilmazkaya1977@gmail.com
[2*] Batman University, Faculty of Engineering and Architecture, Departmant of Computer Engineering, Batman, Turkey, (ORCID: 0000-0003-4325-6922), ramazan.tekin@batman.edu.tr

(First received 28 February 2022 and in final form 16 March 2022)

(**DOI:** 10.31590/ejosat.1080098)

ATIF/REFERENCE: Kaya, Y., & Tekin, R. (2022). Comparison of discretization methods for classifier decision trees and decision rules on medical data sets. *European Journal of Science and Technology*, (35), 275-281.

Abstract

Data sets in real life are given by real numbers in databases. On the other hand, many data mining methods like association rules and induction rules require only discrete attributes. Therefore, in order to use such data mining methods on datasets with continuous characteristics, these features in the dataset should be discretized. The discretization process is reducing the number of values by dividing the range interval of a continuous attribute into certain intervals. In this paper, eight discretization methods are presented with JRip, OneR, J48, and Part classifier algorithms of rules and tress. The applications within the scope of the study include the results obtained as a result of ten-fold cross validation and were carried out on real-life data obtained from the UCI repository. We show that discretization is important step to significantly increase the classification results of these algorithms. Finally, as a result of the study, it was seen that MDL and J48, CAIM and Jrip and Extended Chi and J48 methods gave the highest accuracy for PIMA, WBC and DERMA data sets, respectively.

Keywords: Classification, Continuous attributes, Discretization, Data Mining.

Ayrıklaştırma Yöntemlerinin Karar ağaçlar ve Karar Kuralları Sınıflandırıcılar için Medikal Veri Setleri Üzerinde Karşılaştırılması

Öz

Gerçek hayattaki veri kümeleri, veri tabanlarında reel sayılarla sunulmaktadır. Öte yandan, birliktelik kuralları ve tümevarım kuralları gibi birçok veri madenciliği yöntemi yalnızca ayrık öznitelikler gerektirirler. Bu nedenle sürekli özniteliklere sahip veri kümelerinin ayrık özniteliklere sahip veri kümelerine dönüştürülmesi gerekmektedir. Ayrıklaştırma işlemi, belirli bir sürekli öznitelik verisini aralıklara bölerek değer sayısını azaltmaktır. Bu çalışmada, kural ve ağaç tabanlı JRip, OneR, J48 ve Part sınıflandırıcı algoritmaları ile sekiz ayrıklaştırma yöntemi naliz edilmiştir. Denemeler, UCI veri deposundan alınan gerçek veri setlerinden oluşmakta ve on kat çapraz doğrulamayı sonuçlarını içermektedir. Bu algoritmaların sınıflandırma başarımı önemli ölçüde artırmada ayrıklaştırmanın önemli bir adım araç olduğunu görülmüştür. Son olarak, çalışma sonucunda PIMA, WBC ve DERMA veri setleri için sırasıyla MDL ve J48, CAIM ve Jrip ve Extended Chi ve J48 yöntemlerinin en yüksek doğruluğu verdiği görülmüştür.

Anahtar Kelimeler: Sınıflandırma, Sürekli Öznitelikler, Ayrıklaştırma, Veri Madenciliği.

---

[*] Corresponding Author: ramazan.tekin@batman.edu.tr

1. Introduction

Data mining is a very powerful tool to uncover meaningful information hidden from large databases. Data preparation, which is an important issue in data mining and data warehousing, is one of the techniques frequently used in both cleaning/correcting missing or outliers and data discretization. Several existing data mining techniques, such as association rules and induction rules, cannot handle continuous attributes. For this reason, continuous features are divided into a series of sub-ranges called categories. The process of dividing continuous features into categories is referred to as the discretization process (Chmielewski & Grzymala-Busse, 1996; Jin et al., 2009).

Real data sets can show various characteristics such as big or small size, large or small number of attributes, various data types, and value ranges. In discretization process, several advantages are desirable in order to achieve good classification results, reduction, and simplification(Abraham et al., 2009; Das & Vyas, 2010).

Garcia et al.(Garcia et al., 2012), compared classification successes using various discrimination techniques together with lazy, rule, decision tree, and Bayesian learning classifiers. They state that their study will guide any researcher in terms of which discretization method should be applied to their own data. Ferreira and Figueiredo(Ferreira & Figueiredo, 2012) used a combining of feature selection techniques and unsupervised feature discretization algorithms to examine which of them achieved the best success. Another similar study by Tran et al. (Tran et al., 2017) demonstrated the performance of two-stage approaches by proposing a new approach that combines feature selection and discretization. Tsai and Chen(C.-F. Tsai & Chen, 2019), examined the effects of feature selection and discretization processing order on classification performance. They also focused on the most appropriate associations between feature selection methods and discretization methods.

Hishamuddin et al. in their work(Hishamuddin et al., 2020), applied the fuzzy-based discretization algorithm to increase the accuracy of a group of classifiers. As a result of their studies, they have seen that the fuzzy discretization method gives the highest accuracy with Random Forest. A similar study was carried out by Jun(Jun, 2021). In this study, a new approach named DIMPLED, which includes Evolutionary Algorithm and Multiple sPLits methods, is proposed to increase the performance of decision tree-based classifiers with discretization.

The discrimination methods are divided into various groups as global vs. local, supervised vs. unsupervised, Top-Down vs. Bottom-Up, and static vs. dynamic(Garcia et al., 2012). In this paper we explain the supervised vs. unsupervised methods. Unsupervised methods perform discretization without taking into account class label information (target attributes). The two most familiar unsupervised discretization algorithms, the first is equal-width discretization and the other is equal-frequency discretization(NGUYEN, 1998). Supervised algorithms, such as statistics-based(Kerber, 1992; Liu & Setiono, 1995), entropy-based(U. M. Fayyad & Irani, 1992), and class-attributes interdependency-based algorithms(Kurgan & Cios, 2004) use class information; however, these algorithms do not make use of relations between attributes in the database.

The main focus of this study is determining which one of discretization algorithms supplies more useful discretization for decision trees and decision rules classification methods. In this study, the accuracy performances of discretized data with various discretization methods were compared with the accuracy performances of undiscrete data using Jrip, OneR, J48 and PART classifiers.

For discretization of data sets R system and for classification algorithms Weka system are used. In section 2, information about the discretization methods to be compared in this study is given. Details about the datasets used in the study are presented in the section 3, and the experimental results are presented in the section 4. Finally, the obtained results are compared and discussed in section 5.

2. Material and Method

2.1. Discretization algorithms

In this paper, the following eight discretization methods are presented.

2.1.1. Equal-width discretization (EWD)

It is the most popular and easiest of the unsupervised discretization algorithms. In the Equal-width discretization (EWD) method, the continuous or numeric values of any attribute are divided into equal k intervals between X_{min} and X_{max}. Each cut range equal to $X_{min} + M\ ((X_{max} - X_{min})\ /\ k)$, where M takes on the value from $0, 1, 2, \ldots, (k-1)$ (Li et al., 2010).

2.1.2. Equal-frequency discretization (EFD)

It is another simple an unsupervised discretization method. This algorithm determines the bin boundaries by sorting the data on ascending values of attribute and subsequently divide the data into k equally intervals. In this way, each interval is divided into discrete intervals with an almost equal number of samples. In Equal frequency discretization (EFD) method each interval contains $n\ /\ k$ conjunct values (Li et al., 2010).

2.1.3. Minimum description length (MDL)

Minimum description length (MDL), on offer by Fayyad and Irani, is a supervised hierarchical discretization algorithm(U. Fayyad & Irani, 1993). MDL discretization is a method that uses class information-based entropy to detect categorical boundaries. Decision attribute's information entropy value is the amount of information that would be used for which class an instance belongs (Kotsiantis & Kanellopoulos, 2006). In this approach, an attribute has a range value that initially contains all its values. This interval/range is then recursively divided into smaller sub-intervals according to a specified stopping criterion(U. Fayyad & Irani, 1993).

2.1.4. Chi-merge

Chi-merge is an algorithm that combines adjacent ranges on a bottom-up basis. For this, it is one of the supervised methods because it uses attributes together with class information. It explains the Chi2 criterion to decide whether 2 neighbor intervals are similar to be merged. The relationships between condition and decision attributes values handle by Chi2 is a statistical measure (Kerber, 1992). The Chi2 test is applied for all pairs of split neighbour intervals. Neighbour intervals with the smallest Chi2 value are combined until the specified criterion is met. Here, the merging process is performed based on the Chi2 statistic (such as 0.01, 0.05, and 0.1 levels), which determines the similarities between neighboring intervals. Accordingly, adjacent intervals are merged as long as the Chi2 values are above a certain significance level.

2.1.5. Ameva

An autonomous discretization algorithm (Ameva) approach uses a benchmark based on the Chi2 test. This criterion considers the discrete variable to have the least number of intervals and the least loss of correlation with the target variable (class label) (Gonzalez-Abril et al., 2009).

The Ameva coefficient is describe as follows:

$$Ameva(k) = \frac{\chi^2(k)}{k * (l-1)} \tag{1}$$

here k is the number of intervals, $l \geq 2$ indicates the number of classes. The Ameva rate is calculated from a probability table between the row of the class variable and the column of the discreated ranges (Gonzalez-Abril et al., 2009).

2.1.6. Class-attribute contingency coefficient (CACC)

Class-attribute contingency coefficient (CACC) discretization method is obtained as follows:

$$CACC = \sqrt{\frac{y}{y+M}} \tag{2}$$

$$y = \chi^2 / \log(n)$$

where M and n denote the total number of samples and the number of intervals into which the continuous feature is divided, respectively. The CACC value is calculated from the discreated intervals and class variable. The row matrix specifies the class variable, and the column matrix specifies each discreated range (C.-J. Tsai et al., 2008).

2.1.7. Class-attribute interdependence maximization (CAIM)

The purpose of Class-attribute interdependence maximization (CAIM) method is to incrase the relationship between attribute values with continuous characteristics and class attribute values and to decrase the number of intervals required for discretization. The CAIM algorithm works on the top-down principle based on the greedy approach. The algorithm initially starts with only one interval and iteratively repeats the division using the boundary, which ensures the highest CAIM ratio. One of the critical features that distinguishes the algorithm from other discretization algorithms is that it can automatically determine the number of discrete intervals. The CAIM criterion is calculated between Class (C), discrization (D) and attiribute (F) as in Equation 3:

$$CAIM(C, D|F) = \frac{\sum_{i=1}^{n} \frac{max_i^2}{M_{ir}}}{n} \tag{3}$$

where n and max_i represent the number of intervals and the maximum value in the i. column of the quantile matrix, respectively ($i = 1, 2, \ldots, n$). M_{ir} is the total number of F attribute with continuous values in the range $(d_{r-1}, d_r]$ ($r = 1, 2, \ldots, S$; where S is the class number) (Kurgan & Cios, 2004).

2.1.8. Extended Chi

In order to avoid the inconsistency of the Chi2 algorithm in the extended Chi2 method, a stopping criterion is used for the upper bound (ξ) in each step of the discretization process. The stopping criterion is expressed as (ξ discretized < ξ original) (Su & Hsu, 2005).

2.2. Classification methods

We used classifiers, representatives of different recognition models. These are Jrip, OneR, J48, and Part, which are realized in the Waikato Environment for Knowledge Analysis (Weka) (Hall et al., 2009). The software can be obtained from http://www.cs.waikato.ac.nz/ml/weka/.

2.2.1. JRip (RIPPER)

Jrip is a popular propositional rule learning algorithm based on Repeated Incremental Pruning to Produce Error Reduction (RIPPER). This algorithm generated a detection model composed of rules database that was built to detect new examples (Rajput et al., 2011). For more information see (Holte, 1993).

2.2.2. OneR

The OneR classification method is a simple rule-based method. OneR is the abbreviation of “One Rule”. As the name suggests, the decision tree created in this method is single-level. Rule-based methods create rules based on attributes. Although it is simple, the method, which is very effective, is widely used in machine learning applications(Xu, 2006). For more details see (Holte, 1993).

2.2.3. J48

Another well-known and popular decision-tree-based classification method is J48 proposed by Quinlan(Quinlan, 2014). J48 is a decision tree classification approach. It is a supervised machine learning method that performs estimation of test data based on available attributes. Internal nodes of decision trees specify different attributes. Likewise, the branches (links) between these nodes (attributes) indicate the possible values that the attributes have in the observed samples. In decision trees, extreme values (leaves) indicate the class label. For more details see (Cohen, 1995).

2.2.4. Part

The PART is supervised classification method uses the divide-and-conquer strategy. It recursively creates rules first, then deletes the instances affected by these rules and repeats the process until there are no instances left (Frank & Witten, 1998).

2.3. Performance measures

Classification accuracy is measured using the equation (Menéndez et al., 2010):

$$Accuracy = \frac{Truely\ classified\ objects}{Truely\ classified\ objects} \tag{4}$$

Table 1. All the attributes found in the DERMA dataset.

Clinical Attributes				**Histopathological Attributes**							
ID	**Attribute**	**ID**	**Attribute**	**ID**	**Attribute**	**ID**	**Attribute**	**ID**	**Attribute**	**ID**	**Attribute**
1	Erythema	**7**	Follicular papules	**12**	Melanin incontinence	**23**	Spongiform pustule	**18**	Hyperkeratosis	**29**	Saw-tooth appearance of retes
2	Scaling	**8**	Oral mucosal involvement	**13**	Eosinophils in the infiltrate	**24**	Munro microabcess	**19**	Parakeratosis	**30**	Follicular horn plug
3	Definite borders	**9**	Knee and elbow involvement	**14**	PNL infiltrate	**25**	Focal hypergranulosis	**20**	Clubbing of the rete ridges	**31**	Perifollicular parakeratosis
4	Itching	**10**	Scalp involvement	**15**	Fibrosis of the papillary dermis	**26**	Disappearance of the granular layer	**21**	Elongation of the rete ridges	**32**	Inflammatory monoluclear inflitrate
5	Koebner phenomenon	**11**	Family history (0 or 1)	**16**	Exocytosis	**27**	Vacuolisation and damage of basal layer	**22**	Thinning of the suprapapillary epidermis	**33**	Band-like infiltrate
6	Polygonal papules	**34**	Age (linear)	**17**	Acanthosis	**28**	Spongiosis				

3. Data Sets

The dermatology(*Dermatology Dataset. Available from: Https://Archive.Ics.Uci.Edu/Ml/Datasets/Dermatology*), diabetes(*Pima Indians Diabetes Dataset. Available from: Https://Archive.Ics.Uci.Edu/Ml/Datasets/Diabetes*) and Breast Cancer Wisconsin (Original)(Wolberg & Mangasarian, 1992) datasets are used in our study. All data sets are obtained from UCI Machine learning Repository (UCI-MLR) (Dua & Graff, 2019). In this section, details about the number of instance and features related to the datasets used in the study and their distribution according to classes are presented.

3.1. The dermatology (DERMA) data set

The dermatology (DERMA) dataset consists of 358 samples and 34 attributes remaining after missing observations are deleted. As seen in Table 1, 12 of the features in the data set are clinical and the remaining 22 are Histopathological features. The diseases in this group are lichen planus, seboreic dermatitis, cronic dermatitis, pityriasis rubra pilaris, psoriasis, and pityriasis rosea. Diseases/classes in the DERMA dataset are pityriasis rosea, pityriasis rubra pilaris, lichen planus, psoriasis, cronic dermatitis, and seboreic dermatitis, with sample distributions of 111, 60, 71, 48, 48, and 20, respectively.

3.2. The Pima Indian diabetes (PIMA) data set

PIMA data set contains 768 samples with 8 attributes. All samples were used as there were no missing observations in the data set. The class variable of the dataset indicates whether the person has diabetes or not. It has negative and positive forms. 500 cases are positive and 268 cases are negative. The attributes of the dataset are given in Table 2.

Table 2. Attributes found in the PIMA dataset.

ID	**Attribute**	**ID**	**Attribute**
1	No. of times preg.	**5**	2-h serum insulin
2	Plasma gluc. Conc.	**6**	Body mass index
3	Diast. blood press.	**7**	Diabetes pedigree function
4	Triceps skin fold thickness	**8**	Years of age

3.3. The Wisconsin Breast Cancer (WBC) data set

WBC dataset was collected from the Madison hospital at the University of Wisconsin. The data set consists of 699 samples. 16 samples were excluded from the data set as they contain missing observations. 683 samples were used in the analysis. Information about the data set is given in Table 3. As can be seen, the values of the variables consist of values between 1 and 10. Of the 683 samples, 444 were healthy and 239 were cancer patients.

Table 3. Attributes found in the WBC dataset.

ID	**Attribute**	**ID**	**Attribute**
1	Clump Thickness	**6**	Bare Nuclei
2	Uniformity of Cell Size	**7**	Bland Chromatin
3	Uniformity of Cell Shape	**8**	Normal Nucleoli
4	Marginal Adhesion	**9**	Mitoses
5	Single Epithelial Cell Size		

4. Experimental results

In this study, 2 unsupervised and 6 supervised discretization methods were compared to each other with 4 classification methods. The unsupervised algorithms were equal-width (EWD) and equal-frequency (EFD) and the supervised algorithms were the CAIM, Chi-merge, Extended Chi2, Ameva, CACC, and Fayyad-Irani discretization.

In the following sections, the results of the discretization algorithms on the Pima Indians diabetes, Wisconsin Breast Cancer, and DERMA data sets, a well-known data set from the UCI-MLR, are presented.

For the PIMA data set, the MDL discretization method with a J48 approach showed better results (78.25%) than other discretization and classifier methods (see Table 4). The CACC discretization method with Jrip got a better result (77.9%) than other discretization methods. The accuracy was 76.3% for the PART classifier with the MDL discretization method. Generally, the MDL method gave much better results than other discretization methods. The classification accuracies for the PIMA data set are shown in Figure 1.

Table 4. Classification accuracy of 4 classifiers for the PIMA data set.

Method	Jrip	OneR	J48	Part
EWD	72.6563%	73.5677%	73.8281%	73.4375%
EFD	71.7448%	74.4792%	72.2656%	73.8281%
MDL	77.8646%	74.7396%	**78.2552%**	**76.3021%**
Chi-merge	74.0885%	72.1354%	77.2135%	72.3958%
Ameva	76.5625%	73.5677%	77.2135%	73.4375%
CACC	**77.9948%**	73.5677%	77.9948%	74.0885%
CAIM	75.1302%	**75%**	75.2604%	74.2188%
Extended chi	76.3021%	74.7396%	74.0885%	74.8698%
Without dis.	75.1302%	73.0469%	73.8281%	75.2604%

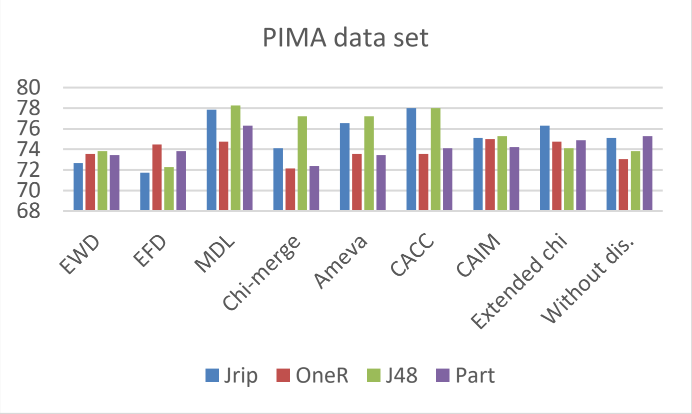

Figure 1. Classification accuracy of 4 classifiers for the PIMA data set.

The classification accuracy of the WBC data set is presented in Table 5. For the WBC data set, the CAIM discretization method with a Jrip approach shows better results than (96.88%) other discretization methods (see Table 5). The extended Chi discretization method with J48 showed better accuracy (96.77%) results than other discretization methods. The the Ameva, CACC, and CAIM discretization methods got better accuracy (96.48%) when compared to other discretization methods. Finally, MDL and Extended Chi discretization methods, with OneR, showed the same classification accuracies (92.67%). The classification accuracies for the WBC data set are shown in Figure 2.

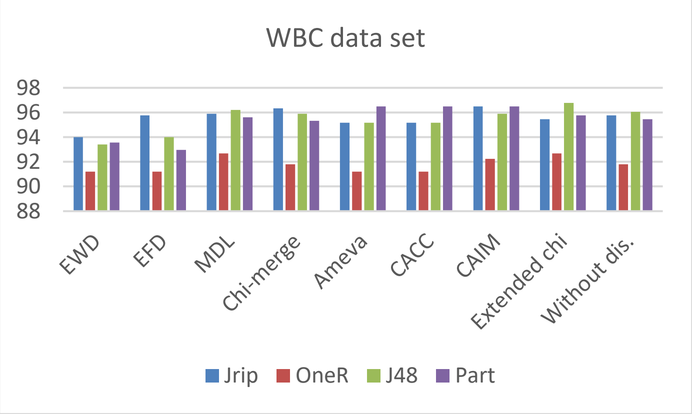

Figure 2. Classification accuracy of 4 classifiers for the WBC data set.

The classification accuracy of the DERMA data set is presented in Table 6. For the DERMA data set, the extended Chi discretization method with J48 approach showed better (96.08%) classification accuracies (see Table 6). The MDL discretization method with Part showed better (95.81%) accuracy results than other discretization methods. The extended Chi2, and MDL discretization methods with the Jrip approach showed better (94.41%) accuracy than other discretization methods. The classification accuracies for the DERMA data set are shown in Figure 3.

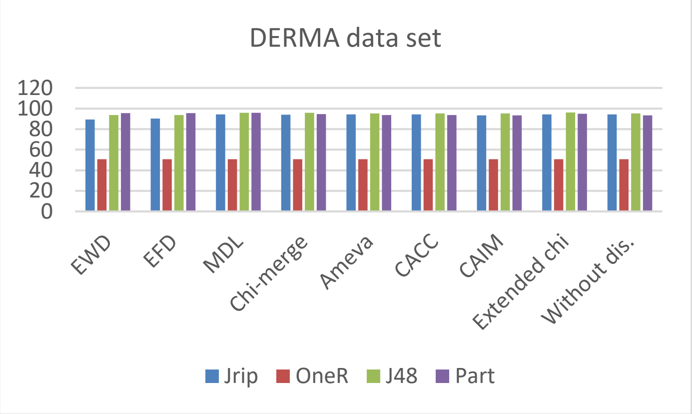

Figure 3. Classification accuracy of 4 classifiers for the DERMA data set.

5. Conclusion

One of the most important operations of data mining preprocessing processes is discretization. Apart from its effect on success, one of the most important reasons for this is that many data mining methods are compatible with attributes with categorical characteristics. Therefore, it is necessary to discretize data with continuous characteristics before it can be processed by methods compatible with such categorical feature sets. The purpose of discretization algorithms can be expressed as dividing the value space into a limited number of categorical values for any feature with continuous characteristics in the data sets.

Table 5. Classification accuracy of 4 classifiers for the WBC data set.

Method	**Jrip**	**OneR**	**J48**	**Part**
EWD	93.9971%	91.2152%	93.4114%	93.5578%
EFD	95.754%	91.2152%	93.9971%	92.9722%
MDL	95.9004%	**92.6794%**	96.1933%	95.6076%
Chi-merge	96.3397%	91.8009%	95.9004%	95.3148%
Ameva	95.1684%	91.2152%	95.1684%	**96.4861%**
CACC	95.1684%	91.2152%	95.1684%	**96.4861%**
CAIM	**96.8861%**	92.2401%	95.9004%	**96.4861%**
Extended chi	95.4612%	**92.6794%**	**96.7789%**	95.754%
Without dis.	95.754%	91.8009%	96.0469%	95.4612%

Table 6. Classification accuracy of 4 classifiers for the DERMA data set.

Method	**Jrip**	**OneR**	**J48**	**Part**
EWD	89.3855%	50.838%	93.5754%	95.5307%
EFD	90.2235%	50.838%	93.5754%	95.5307%
MDL	**94.4134%**	50.838%	95.8101%	**95.8101%**
Chi-merge	93.8547%	50.838%	95.8101%	94.6927%
Ameva	94.1341%	50.838%	95.2514%	93.5754%
CACC	94.1341%	50.838%	95.2514%	93.5754%
CAIM	93.2961%	50.838%	95.2514%	93.2961%
Extended chi	**94.4134%**	50.838%	**96.0894%**	94.9721%
Without dis.	94.1341%	50.838%	95.2514%	93.2961%

In this study, the contribution of discretization algorithms to classification performance in datasets with continuous or both categorical and continuous features is presented comparatively. In the light of the findings, it can be said that discretization in the preprocessing stage greatly increases the performance of the classifier methods. For this purpose, we compared classification results using PIMA diabetes, Wisconsin Breast Cancer and DERMA datasets. It was found that using discretization methods before decision trees and decision rule classifiers achieves better results than using these classifier methods without the discretization process of data. Our study revealed that the discretization methods for Jrip, J48, and PART classifiers lead to an important average increase in accuracy.

Jrip's performance was significantly improved on the PIMA diabetes data set using CACC, on the WBC data set using CAIM, and on the DERMA data set using the extended Chi discretization method. OneR's performance was significantly improved on the PIMA diabetes data set using CAIM, on the WBC data set using the MDL and Extended Chi discretization method, and did not significantly degrade on the DERMA data set. J48's performance was significantly improved on the PIMA diabetes data set using MDL, on the WBC and DERMA data sets using the Extended Chi discretization method. PART's performance was significantly improved on the PIMA diabetes data set using MDL, on the WBC data set using Ameva, CACC, and CAIM, and on the DERMA data sets using the MDL discretization method. The results of the experiments make it clear that discretization techniques can indeed improve the performance of the Jrip, OneR, J48, and PART.

References

Abraham, R., Simha, J. B., & Iyengar, S. S. (2009). Effective Discretization and Hybrid feature selection using Naïve Bayesian classifier for Medical datamining. International Journal of Computational Intelligence Research, 5(2), 116–129.

Chmielewski, M. R., & Grzymala-Busse, J. W. (1996). Global discretization of continuous attributes as preprocessing for machine learning. International Journal of Approximate Reasoning, 15(4), 319–331.

Cohen, W. W. (1995). Fast effective rule induction. In Machine learning proceedings 1995 (pp. 115–123). Elsevier.

Das, K., & Vyas, O. P. (2010). A suitability study of discretization methods for associative classifiers. International Journal of Computer Applications, 5(10), 0975–8887.

Dermatology dataset. Available from: Https://archive.ics.uci.edu/ml/datasets/Dermatology. (n.d.).

Dua, D., & Graff, C. (2019). UCI Machine Learning Repository [http://archive. Ics. Uci. Edu/ml]. Irvine, CA: University of California. School of Information and Computer Science, 25, 27.

Fayyad, U., & Irani, K. (1993). Multi-interval discretization of continuous-valued attributes for classification learning.

Fayyad, U. M., & Irani, K. B. (1992). On the handling of continuous-valued attributes in decision tree generation. Machine Learning, 8(1), 87–102.

Ferreira, A. J., & Figueiredo, M. A. (2012). An unsupervised approach to feature discretization and selection. Pattern Recognition, 45(9), 3048–3060.

Frank, E., & Witten, I. H. (1998). Generating accurate rule sets without global optimization.

Garcia, S., Luengo, J., Sáez, J. A., Lopez, V., & Herrera, F. (2012). A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning. IEEE Transactions on Knowledge and Data Engineering, 25(4), 734–750.

Gonzalez-Abril, L., Cuberos, F. J., Velasco, F., & Ortega, J. A. (2009). Ameva: An autonomous discretization algorithm. Expert Systems with Applications, 36(3), 5327–5332.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. ACM SIGKDD Explorations Newsletter, 11(1), 10–18.

Hishamuddin, M. N. F., Hassan, M. F., & Mokhtar, A. A. (2020). Improving Classification Accuracy of Random Forest Algorithm Using Unsupervised Discretization with Fuzzy Partition and Fuzzy Set Intervals. Proceedings of the 2020 9th International Conference on Software and Computer Applications, 99–104.

Holte, R. C. (1993). Very simple classification rules perform well on most commonly used datasets. Machine Learning, 11(1), 63–90.

Jin, R., Breitbart, Y., & Muoh, C. (2009). Data discretization unification. Knowledge and Information Systems, 19(1), 1–29.

Jun, S. (2021). Evolutionary Algorithm for Improving Decision Tree with Global Discretization in Manufacturing. Sensors, 21(8), 2849.

Kerber, R. (1992). Chimerge: Discretization of numeric attributes. Proceedings of the Tenth National Conference on Artificial Intelligence, 123–128.

Kotsiantis, S., & Kanellopoulos, D. (2006). Discretization techniques: A recent survey. GESTS International Transactions on Computer Science and Engineering, 32(1), 47–58.

Kurgan, L. A., & Cios, K. J. (2004). CAIM discretization algorithm. IEEE Transactions on Knowledge and Data Engineering, 16(2), 145–153.

Li, Y., Liu, L., Bai, X., Cai, H., Ji, W., Guo, D., & Zhu, Y. (2010). Comparative study of discretization methods of microarray data for inferring transcriptional regulatory networks. BMC Bioinformatics, 11(1), 1–6.

Liu, H., & Setiono, R. (1995). Chi2: Feature selection and discretization of numeric attributes. Proceedings of 7th IEEE International Conference on Tools with Artificial Intelligence, 388–391.

Menéndez, L. Á., de Cos Juez, F. J., Lasheras, F. S., & Riesgo, J. Á. (2010). Artificial neural networks applied to cancer detection in a breast screening programme. Mathematical and Computer Modelling, 52(7–8), 983–991.

NGUYEN, H. S. (1998). Discretization methods in data mining. Rough Sets in Knowledge Discovery, 451–482.

Pima Indians Diabetes dataset. Available from: Https://archive.ics.uci.edu/ml/datasets/diabetes. (n.d.).

Quinlan, J. R. (2014). C4.5: Programs for machine learning. Elsevier.

Rajput, A., Aharwal, R. P., Dubey, M., Saxena, S., & Raghuvanshi, M. (2011). J48 and JRIP rules for e-governance data. International Journal of Computer Science and Security (IJCSS), 5(2), 201.

Su, C.-T., & Hsu, J.-H. (2005). An extended chi2 algorithm for discretization of real value attributes. IEEE Transactions on Knowledge and Data Engineering, 17(3), 437–441.

Tran, B., Xue, B., & Zhang, M. (2017). A new representation in PSO for discretization-based feature selection. IEEE Transactions on Cybernetics, 48(6), 1733–1746.

Tsai, C.-F., & Chen, Y.-C. (2019). The optimal combination of feature selection and data discretization: An empirical study. Information Sciences, 505, 282–293.

Tsai, C.-J., Lee, C.-I., & Yang, W.-P. (2008). A discretization algorithm based on class-attribute contingency coefficient. Information Sciences, 178(3), 714–731.

Wolberg, Wi. H., & Mangasarian, O. (1992). Breast cancer wisconsin (original) data set. UCI Machine Learning Repository.

Xu, X. (2006). Adaptive intrusion detection based on machine learning: Feature extraction, classifier construction and sequential pattern prediction. International Journal of Web Services Practices, 2(1–2), 49–58.