Multiclass Classification Methods: A Review

Conference Paper · December 2020

3 authors, including:

Zahoor Mosad
University of Thi-Qar
8 PUBLICATIONS   **36** CITATIONS

Wamidh k. Mutlag
Southern Technical University -Iraq
6 PUBLICATIONS   **42** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   A new system for converting voices of Persian letters into gesture View project

Project   International Journal of Advanced Engineering Technology and Innovative Science (IJAETIS) View project

Multiclass Classification Methods: A Review

Baidaa M Alsafy[1], Zahoor M. Aydam[2], Wamidh K. Mutlag[3]
[1,2] Computer sciences and mathematics college, University of Thi_Qar, Thi_Qar, Iraq.
[3] Al shatrah Technical Institute, Southern Technical University, Iraq
E-mail: [1]baidaaalsafy@utq.edu.iq , [2] zahoormosad@gmail.com, [3]wamid.almuhaysen@stu.edu.iq

ABSTRACT

The methods used for Classification, Recognition, Diagnosis or Clustering with details, equations and features will let us know whether it 'good' or 'bad' in action. Such methods have become increasingly important with the researchers in recent years. A wide range of statistical methods has been applied, but no advance for the researchers which one will be appropriate for their applications. The results from this result will be features that used for classification, recognition, diagnosis or clustering. Feature selection plays a significant role in the classification. In this paper, we collected most important classification methods that may use in researches and compare between these Methods.

Keyword*: Data collection**, Feature extraction; Classification, Classification, Recognition, Diagnosis, Clustering

INTRODUCTION

Classification is a data mining technicality so specifies classes to a set of data in order to help with predictions and analysis. As well as the classification is a function to extract data in a group to base classes or groups. A classification task starts with a data group in whose category tasks do know. The aim of classification is to truly prediction the targeted status for per status in the data and discovers how that set of attributes reaches its conclusion [1].

2. Data collection

There are two methods of collecting data [2]:

1. Quantitative data collection

2. Qualitative data collection

2.1 Quantitative data collection

This procedure is based on taking tools to collect structured data appropriate for the expertise of the categories of predefined responses and random sampling, which results in results that are easy to abstract, compare and published [2].

2.2 Qualitative data collection

Data collection process plays a paramount role in effect assessment by giving useful data to understand methods that are based on observable outcomes and change assessment [3].

The researcher needs to register the data to be useful and accurate through the use of systematic observations, field graphics, photography, and other appropriate means. The methods of collecting data must take into account research ethics. There are many methods of collecting information and a wide range of sources. The following are a few methods of collecting information for research projects [2, 3].

1. Questionnaires

2. Prtfolios

3. Direct observations

4. Inter views

5. Focus group interviews

6. Case-studies

7. Critical incidents

8. Diaries

9. Document and other materials

3. Feature selection Methods

The goal of our study is to compare the performance of classification methods of selection of features

1

when the sizes of training patterns are limited and the number of features is significant. We consider many algorithms to choose common features.

4. Multiclass classification methods

This section briefly illustrates the multiclass classification techniques which are studied in this paper.

4.1 A K-Nearest Neighbors (KNN) Classifier

This kind of classifier has been shown to be efficient in action recognition studies and selects the action that is relative to the merit under question using the Euclidean distance metric in the multidimensional merit space. Many studies are used KNN classifier, Because of its implementational simplicity and elasticity, and the verity that it can let analysis of the classification resolution [4, 5].

KNN is one of the old non-parametric classification techniques. To classify an unbeknown example, distance is measured (using some distance measures such as Euclidean) for example training. The shortest distances of k are specified, and the most visible category in these categories is the category of output class. The k value is typically determined by using the validation group or by using validation [6].

4.2 Support vector machines method

SVMs are among the most strong and effective classification techniques [7]. They are established based the concept of maximizing the margin i.e. maximizing the minimum distance from the separating hyperplane to the nearby example. The essential SVM prop only bilateral classification, but extensions [8] have been suggested to deal with the multiclass classification state too. In these extensions, further parameters and constraints are added to the improvement problem to deal with the segmentation of the various categories [7].

4.3. Decision Tree Method

Decision trees are a strong classification method. It constructs a bilateral classification tree. Every node matches to a bilateral predicate on one feature; one section matches to the positive cases of the predicate and another to the negative cases. So,

every node matches to a sequence of predicates and their rates showing on the descending track from the root to it. Every leaf is categorized through a category. To predict the category label of an input, a track to a leaf from the root is set based on the value of the predicate at every node that is visited. The predicates are selection from highest to lowest through calculating the information profit of every feature that is the expectant decrease in entropy which create by Fragmentation of the samples based on to the feature [9].

4.4. Naive Bayes Method

Naive Bayes is an effective classifier depends on the standard of Maximum A Posteriori (MAP). Specified a trouble with K categories $\{C1, . . . , CK\}$ with so-called previous probabilities $P(C1), . . . , P(CK)$, we can specify the category label c to an obscure example with characters $x = (x1, . . . , xN)$ like $c = argmaxcP(C = ckx1, . . . , xN)$, that is select the category with the maximum a rear probability specific the observed information. This aposterior eventuality can be subedit, by Bayes algorithm, as follows: $P(C = c\\x1, . . . , xN) = P (C=c)P (x1,...,xN \\C=c) /P (x1,...,xN)$. Where the denominator is the same for whole categories, it can be discarded from the comparison. We will count the supposed category conditional probabilities of the characters specified the obtainable categories. This can be completely complicated taking into computation the dependencies among characters. The naive bayes method is to presume category conditional independence i.e. $x1, . . . , xN$ are independent given the category. This facilitates the numerator to be $P(C = c)P(x1\\C = c). . . P(xN \\C = c)$, and then selecting the category c that maximizes this value total the categories $c = 1, . . . , K$. Clearly this method is surely extensible to the state of having more than two categories, and was obvious to implement quite despite of the underlying simplifying suggestion of conditional independence [10].

4.5. C4.5 Classifier

This algorithm is an improvement of the ID3 developed through Quinlan Ross. C4.5 addresses the

absolute and constant attributes of decision tree construction. For dealing with the ongoing features divides C4.5 attribute values for two sections founded on the specified threshold so as to all values are higher than the rest as a child and another. this also deals with the values of the missing features. C4.5 percentage gain is used as a measure to determine the features to create a decision tree [11].

4.6 Genetic algorithm

Genetic algorithm aims at obtaining the best appropriate parameters utilizing the technicality of genetic evolution and stay of the fittest in natural selection. Genetic algorithms permit the removal of false provisions in algorithms and get better the accuracy of text classification.   This is a global international algorithm for potential change, which A natural medium for biological evolution and genetic development and is vastly utilized for its plainness and force. Presently many researchers have utilized that way to improve the method of text classification. The authors show GA for text classification and were utilized In the construction and development of the user form, and others as well provided a steel simulation to improvement the deficiencies of the GA. While the experiential analysis, which display that the amended way is helpful and efficient to text classification [12].

4.7. Neural Network

The ANN, which is commonly called the neural network (NN), is a mathematical pattern or arithmetic pattern depends on the biological neural networks, in another word, a simulation of the biological nervous system. It comprise of a coherent set of artificial neurons processing data utilizing a contact method to calculate. In generality status, ANN is an adaptive method that converts its structure depends on the outside or inside data that streams over the network through the learning stage [13].

4.8 Fuzzy Logic

Fuzzy logic is a decision-making process depended on decree making applied for expert systems and operation rule. Classical Boolean logic varies from Fuzzy logic in this fuzzy logic partial membership is allowed in the group. Classical Boolean logic has on two values in the meaning that the member either belongs or does not belong. Values of one and zero represent a member of the group with one represents absolute membership and zero represents no membership. Fuzzy logic allows partially membership, or group of membership, Which may be any value between zero and one [14].

4.9 K-Means Classifier

K-Means Clustering: Cluster or aggregation is a kind of unsupervised learning technicality that while a start, creates collections Automatic. The items possess which have similar characteristics put in the same group. that algorithm is called k-means in order to it arises distinct clusters. The average of the values in a certain group is the center of that group [15].

4.10 Bayesian Classifier

Bayesian classifiers are statistical classifiers. They can predict the prospects of chapter membership, such as the possibility that a certain number belongs to a particular category. The classification of the theory of default is based on Bayes' theory. Virtualization classifiers have also shown high accuracy and speed while used to big databases [16].

4.11 ID3 Decision Tree Classifier

The ID3 algorithm uses one of the attributes to be the root of the tree and to create a sub node for each possible value of the attributes. If the node is not a sheet, the selection process is repeated and a sub-contract is created through the training data set, otherwise the process of splitting this section of the tree will end. When the decision tree is finished, a new test state is predicted by comparing the values of the new state attributes with the tree nodes starting from the main root down to the prediction value. The ID3 algorithm uses the Entropy code and wins the information gain to calculate and determine the most appropriate attribute of the division [17] : $Entropy(s) = -f_+ \log_2 f_+ - f_- \log_2 f_-$

Entropy is calculated for each value of the values of the studied attribute:

Where:

S: Value of studied character

$f+$: Percentage of positive cases at the value of the studied character (number of positive cases / total number of cases)

$f-$: Number of negative cases at the value of the studied character (number of negative cases / total number of cases)

The general information gain (for the whole dataset) is calculated:

$$\text{Info(main)} = -f(+)/(f+m) * \log 2(f(+)/(f+m)) - f(-)/(f+m) * \log 2(f(-)/(f+m))$$

Information gain is calculated for each value of the studied attribute values:

$$\text{Info(Attrib)} = f(s)/(total) * \text{Entropy(Attrib)}$$

Where:

f: positive ratio ratio f

m: negative character ratio M

S: Value of studied character

is calculated Gain (Attrib) = Info (main) - info (Attrib) , thus selecting the highest profit.

4.12 Linear Discriminant Analysis

LDA is a well-known method used to determine sample groups in a particular set of data. Attempts to divide groups (data classes) with a linear function so that the classes are far apart Possible, but at the same time maintain the distance between individual data samples in a single minimum category. The method assumes that the data in each chapter is distributed normally, but is still applied successfully in many auto-recognition issues [18].

4.13 Quadratic Discriminant Analysis

As the name suggests, QDA is closely related to LDA except for that QDA goes not suppose a normal distribution of data in all classes. in state, the linear function Separating the groups from each other, the function used by QDA is second class and can be mind as LDA Generalization. Because of more complex time constr[19].

4.14. PCL Classifier

PCL is depending on the idea of emerging patterns. It needs to behavior a characteristic selection process before its pattern is established. Those chosen characteristics are then discretized. Then, the essence knowledge models, our emerging models, are originated from the discretized training data. An emerging model is a group of conditions predominatingly including various features, with which most of a category of patterns expression satisfies, but none of other categories patterns satisfy. So, an emerging sample can be considered as a multicharacteristic discriminator. The core spirit of PCL is to utilize top-class multi-characteristic discriminators to create a mass prediction. PCL uses feature sets, does not assume that features are separate; PCL can equip more than a mere prediction or a distance, but numerous interesting basics [20].

5. Comparison and analysis

In this section comparative results for the classification algorithms are to be debated. This comparative analysis explains the aim and limitations of classification algorithms.

6. Discussion

This study was designed to compare the (14) classification methods. The performance classification algorithm accuracy utilized data collections to kappa coefficient from this study, each classification algorithm essential in any one way. The classification efficiency conclusion contains TP values and TN t values. The classification showing calculated through the following method.

$$Accuracy = \frac{TN + TP}{TN + TP + FP + FN}$$

7. Conclusions

This part explains the purpose and limitations of methods are obtained from this review. Table 1

describes the problem of classification algorithm   which is algorithms obtained of this review

Table (1): Comparison of multiclass classification methods

	Description of Algorithms		
NO	Algorithms	Purpose	Limitations
1	KNN	1. It is ease of understanding 2. Training is so quick. 3. strong for noisy training data.	1. Slow Process 2. Classification time is lengthy 3. tricky to obtain the optimal value
2	SVM	1.most active ways in classification, especially common in text classification 2. contrast with ANN, it takes the basic characters of the data. 3.Gives great efficiency in classification	1. More difficult to classify 2. Tricky to explain for resolve, parameter sample. 3.There is a need for a number of key parameters to get the best rating result
3	Decision tree	1. The decision tree has a stellar velocity of learning and velocity of classification. 2. Backing transparence of knowing /classification. 3. Backing multi-	1. A small number of difference in the data can imply so various looking trees. 2. Development of a decision tree may do affect very for unnecessary attributes. ke predicaments of XOR, parity or

		classification.	multiplexer.
4	Naive Bayes	1. To improve the classification execution by eliminating the unrelated options. 2.performance is good 3. The calculation time is short	1.it needs a great number of registers in order to get good results 2.Requires adjustment of its threshold values
5	C4.5	1.appropriat for real -world problems. 2. address lost values 3.Divide the statement high accurately	1. More easy rules. 2.Requires high training samples 3.Unsatisfactory in practice
6	Genetic algorithm	1. it can get the best solutions in a quite short time 2. The random mutation It fairly guaranteed extent that we view a great range of solutions 3. Its coding is simple compared with other algorithms that perform the very process	1.The difficulty of reaching a good way of guidance reflects in fact the work of the algorithm 2. It may not find the optimal resolution for the issue specified in all status. 3.Difficult to choose parameters like population size وa number of generations, etc.

7	ANN	1. In complex the range, it supplies good result 2. best for continued domain 3. The testing operation is speedy	Slow in the training process
8	Fuzzy logic	1. The notions of mathematics used in fuzzy reasoning are easy. Fuzzy logic notions are simple to understand. 2. Fuzzy logic is elastic for any system and is easy to gather more functions without starting from the start and easy management 3. Fuzzy logic afford inaccurate data. 4. Fuzzy logic may be merged with standard control ways. 5. Fuzzy logic is depended on natural language.	Fuzzy logic is a and an appropriate method to set the input range to the output range, but it may not be comfortable in all conditions. If there is a simpler solution, ambiguous logic may not be advisable
9	K-mean	1.Computational complexity is low 2.Simple and easy	1.Depends on a large number of parameters 2.This way does not guarantee the best solution

		implementation 3.Dealing with a total of data on a large scale	3.Failure with sets nonlinear data
10	Bayesian	1. Bayesian classifiers are efficient as decision trees and neural network classifiers. 2. This Very high accuracy 3.shows a very high speed 4.The independence of the saved layer appears 5.Simplifies the arithmetic process	1. Calculations and probabilities using a complex Bayes are often characterized by accuracy and caution must be taken to calculate them correctly 2. The calculate can be NP-hard 3.The quality of the results depends on the quality of the beliefs or the previous model 4. All sections should be calculated for the possibility of one branch calculate
11	ID3	1. It produces s a high precision algorithm of the C4.5 algorithm. 2. It gives omission rate decreased and false alarm rate , Increases detection rate and reduces space consumption 3.ID3 algorithm use nominal attributes for classification with no missing values	1.Search time is long 2.It takes a large amount of memory, which is bigger than the C4.5algorithm to execute a large program

12	LDA	It work for searches for all vectors in the underlying space and find the best discriminate among classes	complex matrix
13	QDA	Close to LDA but no here is no assumption that the covariance of each of the classes is identical, more accurate.	complex matrix
14	PCL	Used convert data from higher- dimensional into a lower-dimensional space	-

References

1. J.Jelina, R.Sasikala, ''**Multi Class Classification Methods on Data Analysis Using Data Mining Techniques''**, IOSR Journal of Engineering (IOSRJEN), ISSN (e): 2250-3021, ISSN (p): 2278-8719, PP 01-04,2018.
2. Male, T, ''**Collecting Quantitative Data**'', In I. Palaiologou, D. Needham and T. Male (Eds). Doing Research in Education: Theory and Practice. London: SAGE, 192-208.],2016.
3. Zubin Austin ,Jane Sutton, '' **Qualitative Research: Data Collection, Analysis, and Management** '', Can J Hosp Pharm , 68(3),pp:226–231,2015.
4. Preece, SJ, Goulermas, JY, Kenney, LPJ and Howard, D,'' **A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data''**, University of Salford,2009.
5. L. Bao and S. S. Intille, **"Activity recognition from user-annotated acceleration data"**,
6. Pervasive Computing, Lecture Notes in Computer Science, vol. 3001, pp. 1-17, 2004.
7. U. Maurer, A. Rowe, A. Smailagic and D. Siewiorek, **"Location and activity recognition**
8. **using eWatch: A wearable sensor platform," ,** Lecture Notes in Computer Science, vol. 3864, pp. 86-102, 2006.
9. Mohamed Aly,'' **Survey on Multiclass Classification Methods''**, November 2005.
10. Yoonkyung Lee, Yi Lin, and Grace Wahba.'' **Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data''.** Journal of the American Statistical Association, 99(465):67–81, March 2004.
11. Tao Li, Chengliang Zhang and Mitsunori Ogihara,' ''**A comparative study of feature selection and multiclass classification**

methods for tissue classification based on gene expression'', Advance Access publication April 15, 2004.

12. Irina Rish. , **''An empirical study of the naive bayes classifier''**, In IJCAI Workshop on Empirical Methods in Artificial Intelligence, 2001.

13. Ian H. Witten, Eibe Frank, & Mark A. Hall, "**Data Mining   Practical Machine Learning Tools and Techniques**", Third Edition, Morgan Kaufmann, 2011.

14. Aurangzeb Khan, Baharum Baharudin, Lam Hong Lee, Khairullah khan, **''A Review of Machine Learning Algorithms for Text-Documents Classification**'', JOURNAL OF ADVANCES IN INFORMATION TECHNOLOGY, VOL. 1, NO. 1, pp:4-20,FEBRUARY 2010.

15. S.Neelamegam, Dr.E.Ramaraj, **'' Classification algorithm in Data mining: An Overview**'', International Journal of P2P Network Trends and Technology (IJPTT) , Volume 3, Issue 5 September to October 2013.

16. J.F. Baldwin, "**Fuzzy logic and fuzzy reasoning**" ,in Fuzzy Reasoning and Its Applications, E.H. Mamdani and B.R. Gaines (eds.), London: Academic Press, 1981.

17. Ayon Dey, " **Machine Learning Algorithms: A Review**", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (3) , pp: 1174-1179,2016.

18. Pang-Ning Tan, Michael Steinbach, & Vipin Kumar, "**Introduction to Data Mining**", Pearson Addison-Wesley, March 2006.

19. Yingying Wang , Yibin Li , Yong Song , Xuewen Rong and Shuaishuai Zhang **,** "**Improvement of ID3 Algorithm Based on Simplified Information Entropy and Coordination Degree**", Algorithms, 10, 124; doi:10.3390/a10040124 ,2017.

20. P Dohnⅉlek, P Gajdoš, T Peterek and V Snⅉšel, "**An Overview of Classification Techniques for Human Activity Recognition**", VIBROENGINEERING PROCEDIA, VOLUME 2, ISSN 2345-0533 ,NOVEMBER 2013.

21. MQ Zhang , " **Identification of protein coding regions in the human genome by quadratic discriminant analysis**", Proceedings of the National Academy of Sciences ,94(2) ,PP:565–8**,**1997**.**

22. Huiqing Liu, Jinyan Li, Limsoon Wong,**'' A Comparative Study on Feature Selection and Classification Methods Using Gene Expression Profiles and Proteomic Patterns''**, Laboratories for Information Technology, 21 Heng Mui Keng Terr, 119613 Singapore, Genome Informatics 13: 51-60 ,2002