International Journal of Computing, Programming and Database Management

E-ISSN: 2707-6644
P-ISSN: 2707-6636
IJCPDM 2021; 2(2): 26-29
Received: 18-05-2021
Accepted: 24-06-2021

Karthik
Department of Computer Science, SDHR College, Tirupati, Andhra Pradesh, India

Pavan Kumar Reddy B
Assistant Professor, Department of Computer Science, SDHR College, Tirupati, Andhra Pradesh, India

Corresponding Author:
Karthik
Department of Computer Science, SDHR College, Tirupati, Andhra Pradesh, India

An experimental approach for prediction of multi-classification using SVM

Karthik and Pavan Kumar Reddy B

DOI: https://doi.org/10.33545/27076636.2021.v2.i2a.30

Abstract
The multiclass classification problem is an important topic in the field of pattern recognition. It involves the task of classifying input instances into one of multiple classes. Since the class overlapping problem exists among multiple classes in most real-world problems, the multiclass classification task is much more complicated and challenging compared to the binary class problem. Classification involves the learning of the mapping function that associates input samples to corresponding target label. There are two major categories of classification problems: Single-label classification and multi-label classification. Traditional binary and multi-class classifications are subcategories of single-label classification. The performance of the developed classifier is evaluated using datasets from binary, multi-class and multi-label problems. The results obtained are compared with state-of-the-art techniques from each of the classification types.

Keywords: Experimental, approach, multi-classification, SVM

1. Introduction

Multiclass classification is a machine learning classification task that consists of more than two classes, or outputs. Machine learning classification is the process of approximating the mapping function that maps the input sample to target class/label [1, 2]. In traditional classification problems, the input samples correspond to only one target label. This type of classification is called single-label classification. Binary classification involves classifying the input data samples into either of two sets based on a specific classification metric. The number of disjoint labels is 2 for binary classification. There are several real-world application problems involving multiple target labels resulting in the development of multi-class classification. Multi-class classification involves classifying the input samples into more than two classes. Character recognition, biometric identification and security, face recognition is some of the application areas of multi-class classification [4, 5].

However, in many real-world applications, the input samples correspond to multiple target labels. This condition of classification, where the input data correspond to a set of class labels instead of one, is called multi-label classification. Multilabel classification has become a rapidly emerging field of machine learning due to the wide range of application domains and the omnipresence of multi-label problems in real world scenarios [6, 8].

So to perform classification tasks, all predictive classification models do not support multi-class classification like Logistic regression, support Vector Machine as those are designed to perform Binary classification and do not support classification tasks more than two classes [3, 7]. In contrast, Decision tree classification, K-nearest neighbour, Naive Bayes Classification and neural network-based models give superior performance for Multi-Class Classification.

Algorithms such as the Decision tree, and KNN were designed for binary classification and do not natively support classification tasks with more than two classes. Instead, heuristic methods can be used to split a multi-class classification problem into multiple binary classification datasets and train a binary classification model each. One approach for using binary classification algorithms for multi-classification problems is to split the multi-class classification dataset into multiple binary classification datasets and fit a binary classification model on each. Two different examples of this approach are the One-vs-Rest and One-vs-One strategies.

2. Multi-Classification

Multi-class classification is those tasks where examples are assigned exactly one of more than two classes.

2.1 One-Vs-rest for multi-class classification

One-vs-rest (OvR for short, also referred to as One-vs-All or OvA) is a heuristic method for using binary classification algorithms for multi-class classification. It involves splitting the multi-class dataset into multiple binary classification problems. A binary classifier is then trained on each binary classification problem and predictions are made using the model that is the most confident.

2.2 One-Vs-One for Multi-Class Classification

One-vs-One (OvO for short) is another heuristic method for using binary classification algorithms for multi-class classification. Like one-vs-rest, one-vs-one splits a multi-class classification dataset into binary classification problems. Unlike one-vs-rest that splits it into one binary dataset for each class, the one-vs-one approach splits the dataset into one dataset for each class versus every other class.

The support vector machine implementation in the scikit-learn is provided by the SVC class and supports the one-vs-one method for multi-class classification problems.

3. Support vector machine

Support Vector Machines (SVM) is a machine learning algorithm that is generally used for classification problems. SVM algorithm is one of the most powerful classification techniques that was successfully applied to many real-world problems [10]. Support Vector Machines are based on the idea of mapping data points to a high dimensional feature space where a separating hyper-plane can be found. The main logic used by SVM for data classification is to drawn optimal hyper-plane which acts as a separator between the two classes. The separator should be chosen like that it gives the maximum margin between the vectors of two classes as shown in figure-1. Due to this reason SVM is also called maximum margin classifier. The vectors near the hyper-plane are called support vectors. This mapping can be carried on by applying the kernel trick which implicitly transforms the input space into another high dimensional feature space. The hyper-plane is computed by maximizing the distance of the closest patterns, i.e., margin maximization, avoiding the problem of over fitting [11].

Consider the two class problem where the classes are linearly separable. Let the dataset D be given as (x_1, y_1), (x_2, y_2)….. (x_n, y_n) ε R^n, where x_i is the set of training tuples with associated class labels, y_i. Each y_i can take one of the two values, either +1 or -1. The data are linearly separable because many number of straight lines can separate the data points into two distinct classes where, in class 1, y = +1 and in class 2, y = -1. The best separating hyperplanes will be the one which have the maximal margin between them. The maximum margin hyperplane will be more accurate in classifying the future data tuples than the smaller margin.

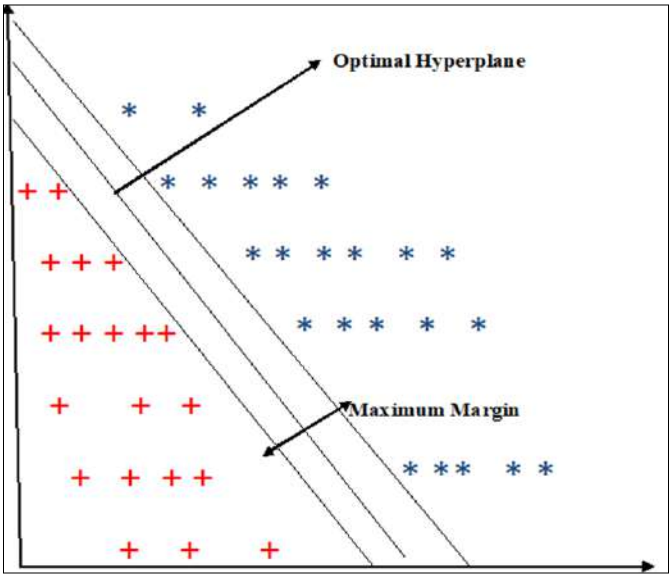

Fig 1: Optimally splitting hyperplane

4. Experimental Results

This section describes the experimental results obtained by applying the proposed multi classification algorithm to a Segment-test data setare taken from the UCI machine learning repository [9]. In the Segment-test dataset, there are 810 records, 20 attributes and 7 class labels are shown in the figure-2. We have used the Python Language to experiment our proposed algorithms. The Python Scikit-learn is a package for data classification, regression, clustering and visualization. The classification models were implemented in Python programming language. The scikit-learn library provides a separate One vs One Classifier class that allows the one-vs-one strategy to be used with any classifier.

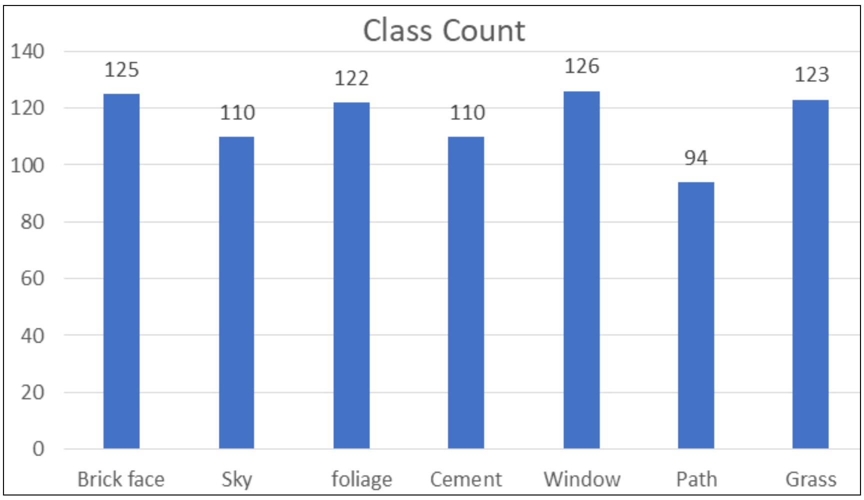

Fig 2: Class-wise distribution of labels

The Segment-test detailed information and summary of statistical analysis as shown in the figure-3.

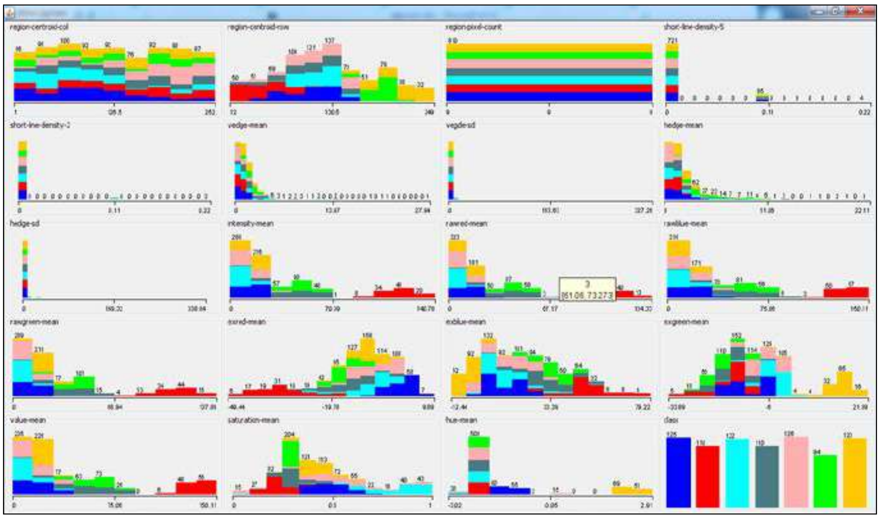

Fig 3: Statistical Summary of Dataset

Table 1: Performance of Mult-classifier

Algorithm	Accuracy	precision	Recall
SVM with One vs One Classifier	95	98	97
SVM	93.5	95	96

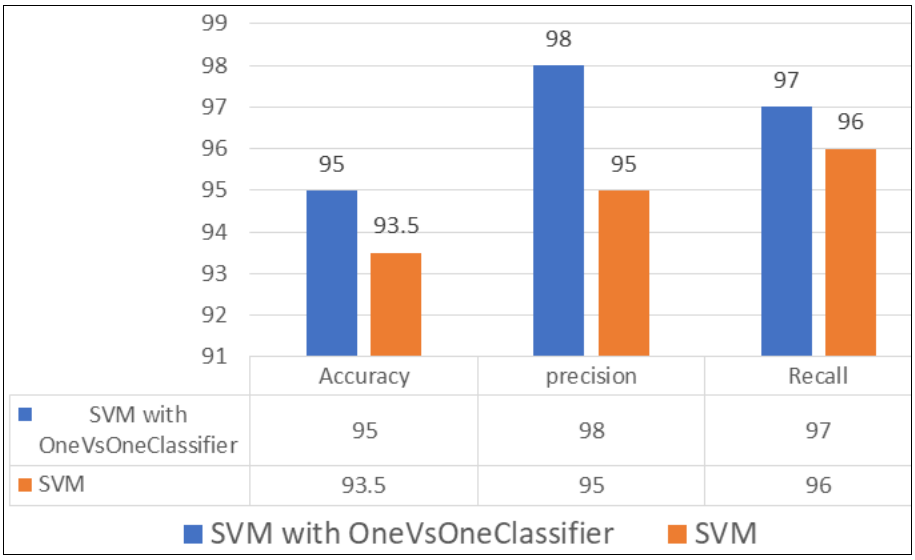

Fig 4: Performance of SVM with One vs One Classifier Multi- Classification

We see in the figure-4, the presentation of the two multi classification order calculations with SVM with One vs One Classifier and SVM based multi classification determination. The accuracy of One vs One Classifier calculation on Segment-test dataset utilizing multi classification has accomplished 95% while SVM based multi classification accuracy has got 93.5.

5. Conclusion

The proposed Segment-test classifier is experimented with nine different datasets of different classification types and the results are compared with state-of-the-art techniques in each type of classification problem. The proposed classifier is evaluated in terms of consistency, speed and performance. The high-speed nature of the proposed classifier makes it suitable for real-time streaming data applications.

6. References

1. Bo G, Xianwu H. SVM multi-class classification, Journal of Data Acquisition & Processing 2006;21:334-339.
2. Tsoumakas G, Katakis I, Vlahavas I. Mining multi-label data, in Data mining and knowledge discovery handbook, ed: Springer 2010, 667-685.
3. Han J, Kamber M. Data Mining Concepts and Techniques. Morgan Kanufmann 2006.
4. Boutell M, Shen X, Luo J, Brown C. Multi-label semantic scene classification, technical report, dept. comp. sci. u. rochester 2003.
5. Pal M. Multiclass approaches for support vector machine-based land cover classification, arXiv preprint arXiv 2008, 0802.2411.

6. Zhang M-L, Zhou Z-H. A review on multi-label learning algorithms, Knowledge and Data Engineering, IEEE Transactions on 2014;26:1819-1837.
7. Tan PN, Steinbach M, Kumar V. Introduction to Data Mining, A: Addision-Wesley 2005.
8. Zhou Z-H, Zhang M-L, Huang S-J, Li Y-F. Multi-instance multi-label learning, Artificial Intelligence 2012;176:2291-2320.
9. UCI machine learning repository. http://archive.ics.uci.edu/ml/
10. Vapnik VN. Statistical learning Theory, John Wiley and Sons, New York, USA 1998.
11. Vapnik VN. The Natural of Statistical Learning Theory, Springer-Verlag, New York, USA 1995.