IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 01-04

Multi Class Classification Methods on Data Analysis Using Data Mining Techniques

J.Jelina, R.Sasikala
M.Phil scholar, Assitant professor, Sankara College of Science and Commerce

Abstract: *Classification problems and their corresponding solving approaches constitute one of the fields of machine learning. Among them classification approaches are categorized into two major types: Two way classification and multi way classification methods are there to analyze the given data. First type of binary or binomial classification is the task of classifying the elements of a given set into two groups (predicting which group each one belongs to) on the basis of a classification rule. In general binary classification is not suitable for all practical problems. Classification problems and their corresponding solving approaches constitute one of the fields of machine learning. Among them classification approaches are categorized into two major types: Two way classification and multi way classification methods are there to analyze the given data. First type of binary or binomial classification is the task of the elements of a given set into two groups on the basis of a classification rule. In this paper we present an overview of the current research being carried out using the data mining techniques to enhance the multi-class classification results. Several major kinds of classification methods including decision tree induction, k-nearest neighbor, neural network, support vector machine and etc., the major aim of this review work is to provide a detailed review of various classification techniques under both two and multi-class in data mining.*

I. Introduction

Classification is a data mining technique that assigns categories to a collection of data in order to aid in more accurate predictions and analysis. Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. For example, a classification model could be used to identify loan applicants as low, medium, or high credit risks. A classification task begins with a data set in which the class assignments are known. The aim of the classification algorithm is to discover how that set of attributes reaches its conclusion .

II. Literature Review

2.1. Review of Two Way Classification Methods

Lu and Harrington [2007] investigated a Gas Chromatography −Differential Mobility Spectrometry (GC−DMS) was tool for analysis of ignitable liquids from fire debris. Headspace Solid-Phase Micro Extraction (SPME) was applied as the pre concentration and sampling. The combined information afforded by gas chromatography and differential mobility spectrometry provided unique two-way patterns for each sample of ignitable liquid. Kesavaraj and Sukumaran [2013] presented the basic classification techniques. Several major kinds of classification method including decision tree induction, Bayesian networks, KNN classifier, the goal of this study is to provide a comprehensive review of different classification techniques in data mining

2.2 Review of Multi-way Classification

A modified Self-Organizing Map (SOM), Kernel-based SOM (KSOM) is introduced to convert the multi-class problems into binary trees, in which the binary decisions are made by SVMs. For consistency between the SOM and SVM the K-SOM utilizes distance measures at the kernel space, not at the input space. Also, by allowing overlaps in the binary decision tree, it overcomes the performance degradation of the tree structure, and shows classification accuracy comparable to those of the popular multi-class SVM approaches with “one-to-one” and “one-to-the others”.

Rosario and Hearst [2005] addressed the problem of multi-way relation classification, applied to identification of the interactions between proteins in bioscience text

III. Results And Discussion

In this section, the results of the classifier are obtained using algorithm (Section 2) on multi class heart disease dataset are discussed in a sequential manner. It contains the records of 303 patients, each of which have 76 features. This database contains 76 attributes, but all published experiments refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by ML researchers to this date.

The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4. Experiments with the Cleveland database have concentrated on simply attempting to distinguish presence (values 1,2,3,4) from absence (value 0).

High recall means that an algorithm returned most of the relevant results. Mathematically, recall is defined as follows:

$$\textbf{Recall} = \frac{TP}{TP+FN} \quad \textbf{(1)}$$

Precision (also called positive predictive value) is the fraction of relevant instances among the retrieved instances. High precision means that an algorithm returned substantially more relevant results than irrelevant ones. Precision is defined as follows:

$$\textbf{Precision} = \frac{TP}{TP+FP} \textbf{(2)}$$

F-measure is defined as the harmonic mean of precision and recall. It is described as follows,

$$\text{F-measure} = 2\,\frac{P.R}{P+R} \quad (3)$$

Accuracy is one metric for evaluating classification models. Informally, accuracy is the fraction of predictions our model got right. For binary classification, accuracy can also be calculated in terms of positives and negatives as follows:

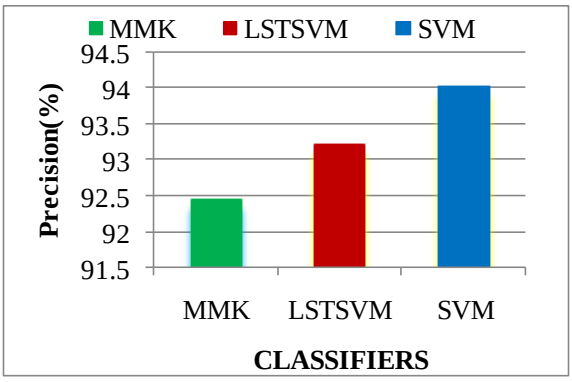
$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

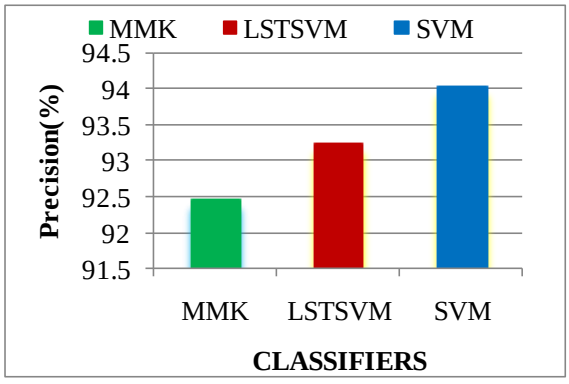
Where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

The performance of the classifier is evaluated using the following standard measures. The results are discussed in table 2.

Table 2. Performance metric comparison

Classifiers	Metrics (%)			
	Precision	Recall	F-measure	Accuracy
MMK	92.45283	94.96124031	93.70703525	89.10891089
LSTSVM	93.233083	95.01915709	94.1261199	89.70099668
SVM	94.029851	96.55172414	95.29078744	91.74917492





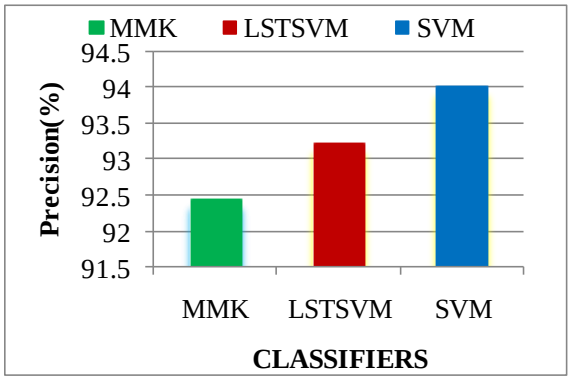


Figure 1. Precision results comparison vs. classifiers

Figure 1 shows the performance comparison results of the precision metrics with respect to three different classifiers such as MMK, SVM, and LSTSVM. From the results it concludes that the proposed SVM classifier produces higher precision results of 94.02%, whereas other methods such as LSTSVM, and MMK classifiers produces only 93.23 % and 92.45% values respectively.

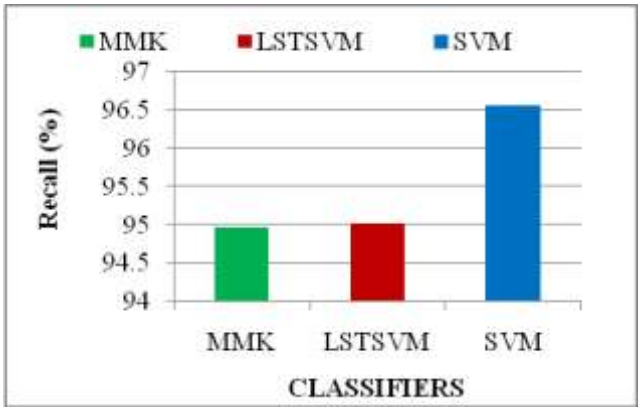


Figure 2. Recall results comparison vs. classifiers

Figure 2 shows the performance comparison results of the recall metrics with respect to three different classifiers such as MMK, SVM, and LSTSVM. From the results it concludes that the proposed SVM classifier produces higher recall results of 96.55%, whereas other methods such as LSTSVM and MMK classifiers produces only 95.01% and 94.96% values respectively.

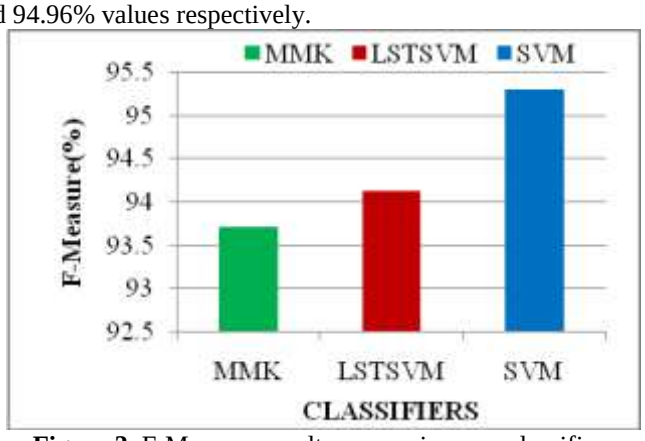


Figure 3. F-Measure results comparison vs. classifiers

Figure 3 shows the performance comparison results of the f-measure metrics with respect to three different classifiers such as MMK, SVM, and LSTSVM. From the results it concludes that the proposed SVM classifier produces higher F-measure results of 95.29%, whereas other methods such as LSTSVM , and MMK classifiers produces only 94.12% and 93.70% values respectively

References

[1]. Amancio, D.R., Comin, C.H., Casanova, D., Travieso, G., Bruno, O.M., Rodrigues, F.A. and da Fontoura Costa, L., 2014. A systematic comparison of supervised classifiers. PloS one, 9(4), p.e94137.

[2]. Bosch, A., Zisserman, A. and Munoz, X., 2007, Image classification using random forests and ferns. IEEE 11th International Conference on Computer Vision, pp. 1-8.

[3]. Bourel, M. and Segura, A.M., 2018. Multiclass classification methods in ecology. Ecological Indicators, 85, pp.1012-1021.

[4]. Cao, M.D. and Zukerman, I., 2012. Experimental evaluation of a lexicon-and corpus-based ensemble for multi-way sentiment analysis. In Proceedings of the Australasian Language Technology Association Workshop ,pp. 52-60

[5]. Cheong, S., Oh, S.H. and Lee, S.Y., 2004. Support vector machines with binary tree architecture for multi-class classification. Neural Information Processing-Letters and Reviews, 2(3), pp.47-51.

[6]. Doan, T.N., Do, T.N. and Poulet, F., 2013, Multi-way classification for large scale visual object dataset. 11th International Workshop on Content-Based Multimedia Indexing (CBMI), pp. 185-190.

[7]. Doan, T.N., Do, T.N. and Poulet, F., 2013, Multi-way classification for large scale visual object dataset. International Workshop on multimedia indexing