OVO Weighted Voting for Multi-Class Imbalanced Classification Having Distance as Weight

Ch.Sarada , M.Sathya Devi

Abstract- Multi-Class imbalanced classification happens in many applications. Most existing solutions use decomposition strategies to address Multi-Class imbalanced classification problem. The common decomposition strategies used are 'One-versus-One' (OVO) and 'One-versus-All'(OVA) binarization techniques. In almost all existing 'One-versus-One' decomposition strategy applications, classification decision has been made based on simple voting strategy, which can lead to wrong conclusions. So, it is necessary to look for alternative. In this study, our aim is to examine empirically the 'One-versus-One weighted voting' decomposition strategy to solve Multi-Class imbalanced classification problem. The proposed methodology consists of 4 steps: In the first step, the original dataset is decomposed into subsets with respect to 'One-versus-One' binarization technique. In the second step, resampling algorithm is applied against each subset to get balanced data. In the third step, SVM learning method is used to construct the binary classifier. In the fourth step, to achieve classification goal, weighted aggregation scheme is applied. Detailed experimental study is epitomized, supported by statistical analysis tool R.

Key words: Multi-Class Imbalanced classification, One-versus-One, One-versus-All, One-versus-One voting, One-versus-One weighted voting, Random over-sampling, Random Under-Sampling, SMOTE, SVM.

1 INTRODUCTION

Imbalanced classification problem typically refers to a problem where one or more classes (usually, the ones that are of interest), are under-represented in the data-set [8]. In such cases, standard classifiers tend to be overwhelmed by the large classes and ignore the small ones [8] leading to a deceptive result. Handling imbalanced data is considered as a challenging task under data mining area [19]. Several solutions have been proposed to deal with two-class imbalanced classification problem [14,16].

Solutions proposed for binary class imbalanced classification problems are not suitable [6,9] for Multi-Class imbalanced classification. One conventional approach to deal with a Multi-Class imbalanced classification is to use binarization techniques, where the original problem is decomposed into several easier binary problems [5].

Commonly used decomposition techniques are 'One-versus-One' [10] and 'One-versus-All' [7]. Binary classification problem in 'One-versus-One' is simpler than "One-versus-All" because in 'One-versus-All' multimodality is usually present, which makes classifier training harder [30]. Results of 'One-versus-One' scheme is determined through aggregation model. With 'One-versus-One' scheme, 'voting strategy' and 'Weighted voting strategy' are the commonly used aggregation models.

There is a limitation with 'One-versus-One voting' decomposition strategy [30]. In 'One-versus-One voting' strategy, voting could be tied [30]: For instance, when number of classes are 3, in 'One-versus-One voting' the 3 class problem is divided into (1,2)th, (2,3)th and (1,3)th binary class problems. Let f1,2(x), f2,3(x) and f1,3(x) be learned decision functions for (1,2)th ,(2,3)th and (1,3)th binary classification problems respectively. Then, the test sample x maybe classified as below.

$f1,2(x) \geq 0$ => Vote for class 1.
$f2,3(x) \geq 0$ => Vote for class 2.
$f1,3(x) < 0$ => Vote for class 3.

The above prediction illustrates that all the classes get equal number of votes. So, decision can't be taken on prediction class as such. This is called tie vote situation. In such a situation, 'weighted voting' per the value of fy,y'(x) could be the practical option. In the 'weighted voting' strategy, each binary classifier assign weight for each vote. The class with the largest sum value is the final output class.

In this study, to address the Multi-Class imbalanced classification, 'One-versus-One weighted voting' scheme is developed in combination with resampling techniques such as random under-sampling, random over-sampling and smote.

This paper aims to develop experimental study on the Multi-Class imbalanced classification using 'One-versus-One weighted voting' decomposition method. The rest of the paper is organized as follows: Section 2 provides a brief outline of the research work carried out in the area. Section 3 provides proposed system architecture. Section 4 gives detailed description of datasets and present the experimental results obtained and analysis. Finally, the conclusions are briefly summarized in Section 5.

2 RELATED WORK

This section first introduces the Multi-Class imbalanced classification problem. Then, the solutions for class imbalanced classification problems are reviewed briefly. Finally, the OVO decomposition strategy for dealing with Multi-Class imbalanced classification problems are described.

2.1 Multi-Class Imbalanced Classification Problem

Class imbalance problem is one fundamental problem in data classification. In the binary class imbalance problem number of instances of one class vastly be more than the other. The Multi-Class Imbalanced classification problem is an extension of the traditional binary class imbalanced classification problem where a dataset consists of three or more classes. An imbalance is said to exist in the Multi-Class imbalanced classification problem, when one or more classes severely outnumber the other classes.

2.2 Solutions for Imbalanced Class Problem

The problem with imbalanced datasets is that the standard classification learning algorithms often ignore the minority class and therefore there is a higher misclassification rate for the minority class instances. Therefore, multiple solutions have been proposed to deal with this problem. They fall into four major groups:

2.2.1 Data sampling: Training instances are resampled in such a way that in the resultant dataset, all classes are equally represented and that allow classifiers to perform in a similar manner to standard classification [13,17,18,19].

2.2.2 Algorithmic modification: This procedure is oriented towards the adaptation of base learning methods which will take imbalance into consideration [26,27,28,29].

2.2.3 Cost-sensitive learning: This type of solution incorporates approaches at the data level, at the algorithmic level, or at both levels combined, considering higher costs for the misclassification of examples of the positive class with respect to the negative class, and therefore, trying to minimize higher cost errors [24,25].

2.2.4 Ensemble learning: Ensemble learning combines a series of learning models with an aim of creating an improved composite classification model. An ensemble tends to be more accurate than its base classifiers [23].

2.3 One-Versus-One Decomposition Strategy to Deal with Multi-Class Imbalanced Classification Problem

Decomposition strategy is considered as the effective scheme to handle the Multi-Class imbalanced classification problems. An easy way to undertake a Multi-Class imbalanced classification problem is to use binarization techniques, where the original problem is decomposed into several easier binary problems [22]. The 'One-versus-One' and the 'One-versus-All' are the two most popular strategies for Multi-Class imbalanced classification. Recent studies clearly proved superiority of OVO over OVA approach [11,12].

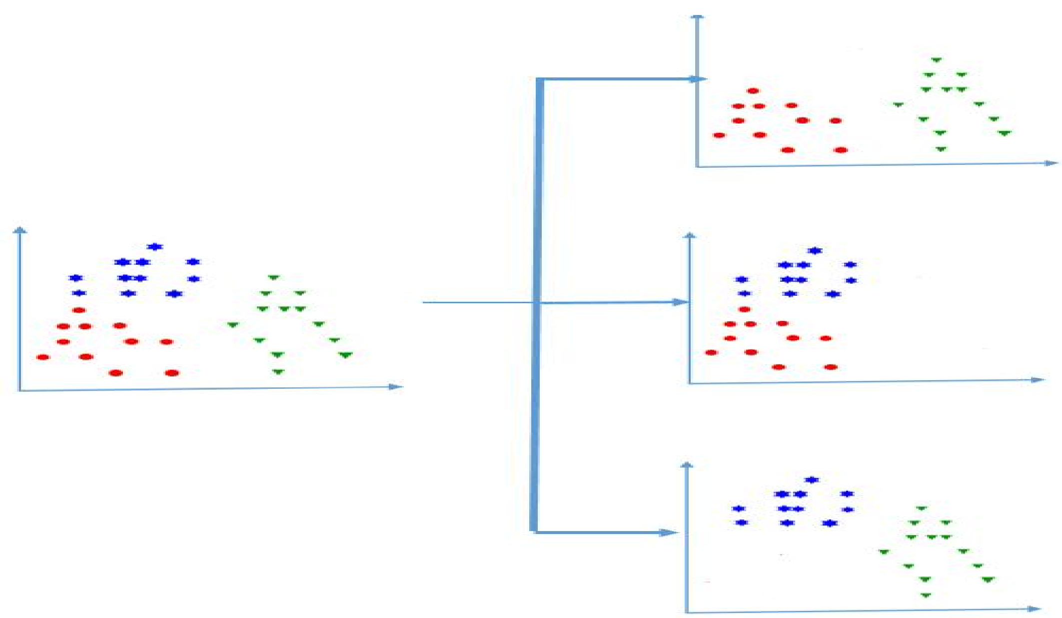

Fig.1. An example of decomposition of 3 class problem into 'One-versus-One' three binary class problem

OVO decomposition states that a 'm' class problem is divided into 'm(m − 1)/2' binary problems. Each problem is allotted to a binary classifier which is responsible for distinguishing between the pair of classes. An example of binarization technique to decompose the Multi-Class problem into 3 binary class problems is shown in Fig. 1.

To predict the class of new pattern, the pattern is presented to every binary classifier. The prediction decision of new pattern is made by aggregating the decisions from learned binary models. There are two aggregation models to finalize the class of new instance such as 'voting strategy' and 'weighted voting strategy'.

Voting strategy is the simplest method to compute the output from OVO classifiers. It is also called Max-Wins rule. In voting strategy, each binary classifier gives a vote for the predicted class. Then, the votes received by each class are counted and the class with the largest number of votes is selected as the final output. To be more precise the binary classification problem assigns label '+1' to samples that belong to the class and '-1' to samples that doesn't belong to the class. Let fy,y'(x)be a learning function for the (y,y')th binary classification.

The learning function fy,y'(x) can be defined as below:

fy,y'(x) ≥ 0=> Vote for class y.

fy,y'(x) < 0 => Vote for class y'.

The test sample 'x' is classified into class that gathers the highest votes. Though voting strategy is easy, it has a constraint that while predicting the new pattern, each class could get same number of votes. This situation leads to a tie in voting. In those conditions, weighted voting strategy is preferred approach. Hence, in this study the weighted voting strategy is proposed.

3 SYSTEM ARCHITECTURE

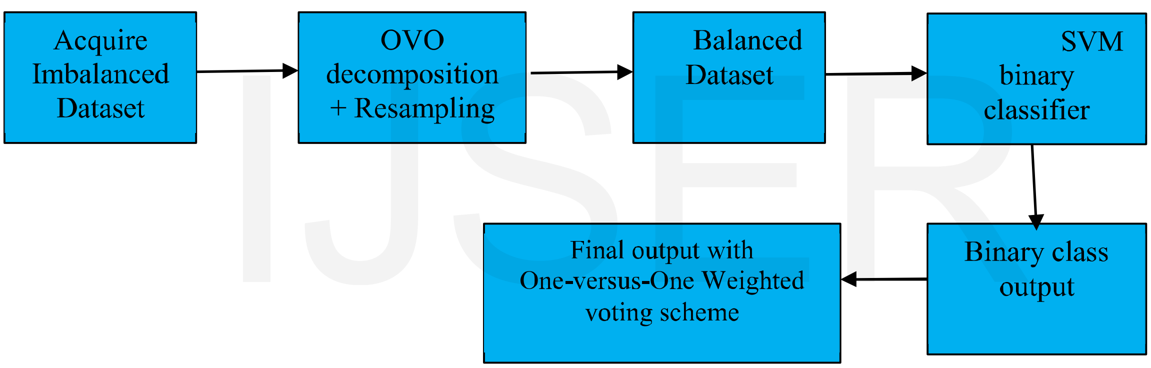

Fig.2. Proposed System Architecture

In this section, architecture of proposed system is explained. Fig. 2 shows the working of proposed system. Resampling algorithm along with 'One-versus-One' technique is used to balance the imbalanced data. Super Vector Machine(SVM) classification is used for building binary classifiers. Finally, 'One-versus-One' weighted voting scheme is used for classification purpose.

3.1 Resampling Techniques

In the proposed system, following Resampling techniques are used to generate balanced data sets.

3.1.1 Random over-sampling[ROS]:

Random over-sampling(ROS)[13] randomly duplicates the minority class samples to balance the class distribution. One pitfall with over-sampling is that it may lead to overfitting as it makes exact copies of the minority samples.

3.1.2 Random Under-Sampling[RUS]:

RUS [17] is the under-sampling technique that randomly discards the majority class instances to balance the class distribution. However, random under-sampling may discard potential useful majority samples.

3.1.3 SMOTE:

SMOTE is an intelligent over-sampling approach proposed by Chawla et al. [8]. In SMOTE, minority class is over-sampled by creating "synthetic" examples rather than by over-sampling using replacement. Unlike ROS duplicates the minority examples, SMOTE produces synthetic minority class examples by *k* nearest neighbors, augmented with randomized interpolation. However, the noise might be included in the synthetic minority class examples.

Fig. 3 Demonstrate an example of sampling methods applied over a decomposed 3 class imbalanced dataset.

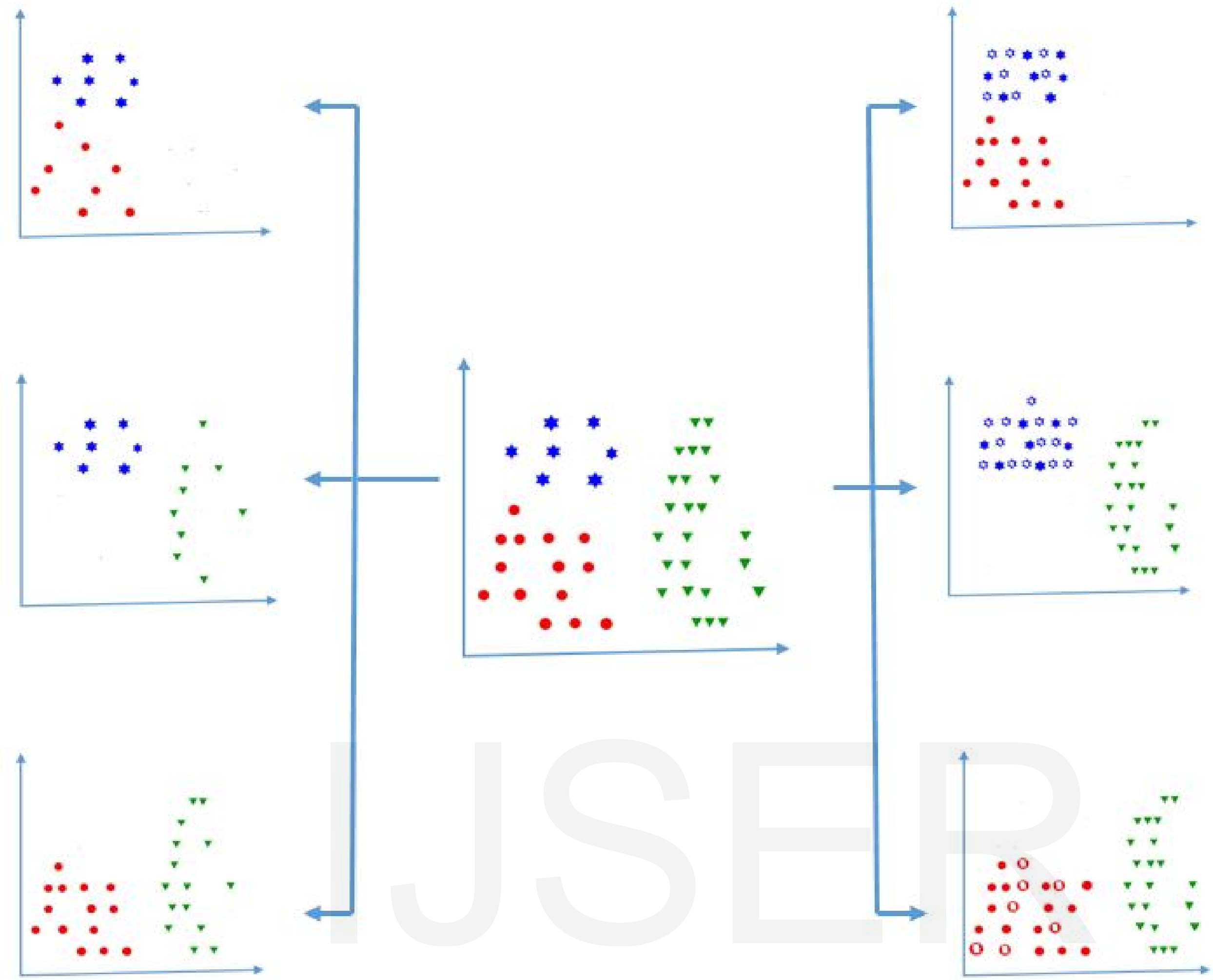

Fig. 3. An example of sampling methods applied over a decomposed 3 class imbalanced dataset. (Left) Under-sampling applied to each binary problem. (Right) Over-sampling applied to each binary problem.

3.2. SVM for Classification:

A Support Vector Machine is a supervised machine learning algorithm which can be used for both classification and regression problems. Kernel trick technique is followed in SVM to transform the data. Optimal boundary is found between possible output based on these transformations. The goal of SVM is to find the optimal hyperplane. The optimal hyperplane is the one which has the biggest margin. Here the aim is to choose the hyperplane which is as far as possible from the data points of each category. The advantage of SVM is that the classification result is highly accurate and the problem of overfitting is less likely to occur when compared to other methods [32]. Fig.4 illustrates Support vectors with optimal hyperplane. In this study, point to hyperplane distance is taken as the weight. Weight is directly proportional to the distance from data point to hyperplane.

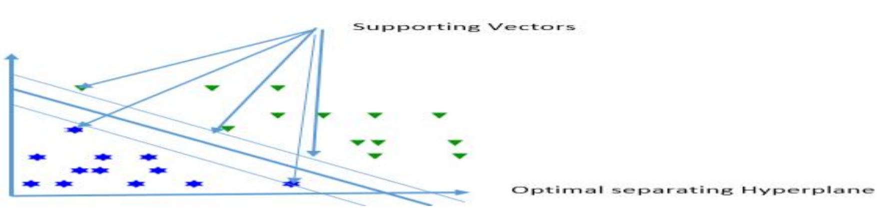

Fig.4. Support Vectors with optimal separating hyperplane

4 EXPERIMENTAL FRAMEWORK

In this section, the details of a couple of real-world problems having Multi-Class imbalanced data, performance measures and the experimental study results are shown.

4.1. Datasets and Parameters

For experimentation, datasets from UCI database repository are used. Table I summarizes the details of selected datasets including number of examples (#EX), number of attributes(#ATTR), Class Distribution and number of classes (#CL), and the imbalance rate (IR)

Table I

Summary of Data sets used in the experimental study

DATASET	#Ex	#Atts	#Cl	#DC	IR
new-thyroid	215	5	3	150/30/35	5.03
balance-scale	625	5	3	288/49/288	5.88
car	1728	7	4	1210/384/69/65	18.62
cmc	1473	10	3	629/233/512	2.69
hayes-roth	132(training)+ 28(testing)	6	6	51/51/30	1.7

4.2. Performance Measures The valuation is a crucial factor in measuring the classification performance and guiding the classifier construction. Large number of performance measures exist in imbalanced classification like precision, sensitivity, G-mean, F-measures and AUC. However, for this study, average accuracy rate is used as the performance measure.

The average accuracy is computed as follows:

$$\text{AveAcc} = \frac{1}{m}\sum_{i=1}^{m} \text{TRP}_i,$$

where m is the number of classes and TRPi stands for the True Positive Rate [33] of the ith class.

True positive rate of a class is defined as percentage of positive samples which are correctly classified samples and it is given as:

$$\text{True Positive Rate} = \frac{\text{True positive samples}}{\text{True positive samples} + \text{False negative samples}}$$

4.3 Experimental Study

To do an experiment on the datasets described in Table-I, half of the samples of each class are used as training data and remining half are used for testing. Results of testing data for all datasets described in Table I are summarized in Table II and number of tie vote samples in OVO voting are summarized in Table III. To make the results more intuitive, results of Multi-Class SVM, 'One-versus-One voting' scheme is compared with proposed 'One-versus-One weighted voting' strategy results. Experimental study is done using R programming tool. The ovun.sample() method of "ROSE" is used for random under-sampling and random over-sampling. The smote () method of "e1071" is used for synthesized over-sampling.From Table III, it can clearly be seen that in 'One-Versus-One voting' scheme tie vote problem occurs. So, the proposed system, 'OVO weighted voting' for Multi-Class Imbalanced Classification having distance as weight' certainly be one solution.

Table II results shows an accuracy improvement with proposed system (Improvements are highlighted). In 'One-versus-One voting' scheme, by considering tie voting the true positive rate may be calculated as:

$$\text{True Positive Rate}_{\text{(one-versus-one voting)}} = \frac{\text{True Positive samples}}{\text{True Positive samples} + \text{False Negatives} + \text{Number of tie samples}}$$

'One-versus-One voting' strategy accuracy is calculated using True Positive Rate $_{\text{(one-versus-one voting)}}$.

Table II

Comparison of accuracy [%] results of Multi-class SVM, OVO voting with resampling and OVO weighted voting with resampling. Standard SVM binary classifier used for both 'One-versus-One voting' and 'One-versus-One weighted voting'.

DATASET	SVM OUTPUT WITHOUT RESAMPLING	One-Versus-One			weighted One-Versus-One		
		RUS	ROS	SMOTE	RUS	ROS	SMOTE
new-thyroid	91.63	95.11	97.33	95.88	97.78	**99.11**	97.78
balance-scale	65.74	74.76	82.17	76.62	75.69	72.68	72.68
car	86.39	83.64	86.31	84.63	81.21	81.86	82.43
cmc	51.14	40.5	52.46	49.57	46.93	**55.22**	52.3
hayes-roth	80.03	82.78	85.16	85.16	84.98	**85.16**	**85.16**

Table III

Number of tie vote samples in One-versus-One voting

DATA SET	RESAMPLING TYPE		
	RUS	ROS	SMOTE
new-thyroid	2		1
balance-scale			
car			
cmc	13	16	22
hayes-roth	2		

In this paper, 'One-versus-One weighted voting' strategy is presented to realize Multi-Class imbalanced learning. Sampling techniques are used to balance the data and Support Vector Machine(SVM) is used for binary classification. The performance of the proposed system is tested in terms of average accuracy. Experimental study is carried out using UCI repository [34]. The results clearly exposed that 'One-versus-One weighted voting' strategy is one practical solution for Multi-Class problem.

In this study, point to plane distance is directly taken as the weight (weight is proportional to the distance between the point and the plane). Further, to improve the performance of Multi-Class imbalanced classification, OVO weighted scheme could be developed using ensemble approach. Moreover, the OVO weighted scheme can be extended to Bigdata using Hadoop environment.

6 ACKNOWLEDGEMENT

The authors would like to expression their sincere gratitude to CVR College of Engineering for their encouragement. The authors like to say thanks to all their colleagues for their support.

7 REFERENCES

[1]. Phua, Chun Wei Clifton; Alahakoon, Damminda; Lee, Vincent Cheng-Siong. Minority Report in Fraud Detection: Classification of Skewed Data. SIGKDD Explorations: newsletter of the Special Interest Group (SIG) on Knowledge Discovery and Data Mining, Vol. 6, No. 1, 2004, p. 50 - 59.

[2]. Shankar, UmaRodda, Sireesha Erothi, Rao,Class Imbalance Problem in the Network Intrusion Detection Systems, ICEEOT-2016,pp:2685-2688.

[3]. Sathyan Munirathinam and Balakrishnan Ramadoss ,Predictive Models for Equipment Fault Detection in the Semiconductor Manufacturing Process , IACSIT International Journal of Engineering and Technology.

[4] . M.H.B. Abidine, B. Fergani, A new multi-class wsvm classification to imbalanced human activity dataset, J.Comput. 9 (7) (2014) 1560–1565.

[5]. Problem T. Ryan Hoens1 , Qi Qian2 , Nitesh V. Chawla1 , and Zhi-Hua Zhou2,Building Decision Trees for the Multi-class Imbalance .

[6]. G.E.A.P.A. Batista, R.C. Prati, M.C. Monard, A study of the behaviour of several methods for balancing machine

learning training data, SIGKDD Explor. Newsl. 6 (1) (2004) 20–29
[7]. L. Cerf, D. Gay, N. Selmaoui-Folcher, B. Cremilleux, J.F. Boulicaut, Parameter free classification in multi-class imbalanced data sets, Data Knowl. Eng. 87(2013) 109–129.
[8]. N.V. Chawla, N. Japkowicz, A. Kolcz, Editorial: special issue on learning from imbalanced data sets, SIGKDD Explor. Newsl. 6 (1) (2004) 1–6.
[9]. A. Estabrooks, T. Jo, N. Japkowicz, A multiple resampling method for learning from imbalanced data sets, Comput. Intell. 20 (1) (2004) 18–36
[10]. M. Galar, A. Fernandez, E. Barrenechea, F. Herrera, Empowering difficult classes with a similarity-based aggregation in multi-class classification problems, Inform. Sci. 264 (2014) 135–157.
[11]. M. Galar, A. FernÈíndez, E. Barrenechea, H. Bustince, F. Herrera, An overview of ensemble methods for binary classifiers in multi-class problems: Experimen- tal study on one-vs-one and one-vs-all schemes, Pattern Recog. 44 (8) (2011) 1761–1776.
[12]. M. Galar, A. FernÈíndez, E. Barrenechea, H. Bustince, F. Herrera, Dynamic classifier selection for one-vs-one strategy: avoiding non-competent classifiers, Pattern Recog. 46 (12) (2013) 3412–3424.
[13]. S. Garcia, J. Derrac, I. Triguero, C.J. Carmona, F. Herrera, Evolutionary-based selection of generalized instances for imbalanced classification, Knowl. BasedSyst. 25 (1) (2012) 3–12.
[14]. H. Haibo, E.A. Garcia, Learning from imbalanced data, IEEE Trans. Knowl. Data Eng. 21 (9) (2009) 1263–1284.
[15] B. Krawczyk, M. Galar, Ł. Jele n´ , F. Herrera, Evolutionary undersampling boost- 846 ing for imbalanced classification of breast cancer malignancy, Appl. Soft Comput. (2015). http://dx.doi.org/10.1016/j.asoc.2015.08.060 [16]. V. Lopez, A. Fernandez, S. Garcia, V. Palade, F. Herrera, An insight into classification with imbalanced data: empirical results and current trends on using data intrinsic characteristics, Inform. Sci. (2013) 113–141.
[17]. M.A. Tahir, J. Kittler, F. Yan, Inverse random under sampling for class imbalance problem and its application to multi-label classification, Pattern Recog. 45 (10) (2012) 3738–3750.
[18]. Joseph Prusa, Taghi M. Khoshgoftaar, David J. Dittman, Amri Napolitano , Using Random Undersampling to Alleviate Class Imbalance on Tweet Sentiment Data ,2015 IEEE 16th International Conference on Information Reuse and Integration.
[19]. Q. Yang, X. Wu, 10 Challenging problems in data mining research, Int. J. Inform. Tech. Decis. 5 (4) (2006) 597–604.
[20](G.E.A.P.A. Batista, R.C. Prati, M.C. Monard, A study of the behaviour of several methods for balancing machine learning training data, SIGKDD Explorations 6 (1) (2004) 20–29. doi: 10.1145/1007730.1007735.
[21]. Z. Zhang et al., Empowering one-vs-one decomposition with ensemble learning for multi-class imbalanceddata, Knowledge-Based Systems (2016), http://dx.doi.org/10.1016/j.knosys.2016.05.048.
[22]. Aggregation schemes for binarization techniques. Methods' description Mikel Galara , Alberto Fern´andezb , Edurne Barrenecheaa , Humberto Bustincea , Francisco Herrerac.
[23]. M. Galar, A. FernÈíndez, E. Barrenechea, H. Bustince, F. Herrera, An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes, Pattern Recog. 44 (8) (2011) 1761–1776.
[24]. Nguyen Thai-Nghe, Zeno Gantner, and Lars Schmidt-Thieme, Cost-Sensitive Learning Methods for Imbalanced Data, DOI: 10.1109/IJCNN.2010.5596486
[25]. W. Penar, M. Wo ´zniak, Cost-sensitive methods of constructing hierarchical classifiers, Expert Syst. 27 (3) (2010) 146–155.
[26]. D.A. Cieslak, T.R. Hoens, N.V. Chawla, W.P. Kegelmeyer, Hellinger distance decision trees are robust and skew-insensitive, Data Min. Knowl. Discov. 24 (1) (2012) 136–158.
[27] . W.M. Czarnecki, J. Tabor, Multithreshold entropy linear classifier: theory and applications, Expert Syst. Appl. 42 (13) (2015) 5591–5606.
[28]. Y.C. Tang, Y.Q. Zhang, N.V. Chawla, S. Krasser, Svms modeling 891 for highly imbalanced classification, IEEE Trans. Syst. Man Cybernetics-Part B 39 (1) (2009) 281–288.
[29] . B. Zadrozny, C. Elkan, Learning and making decisions when costs and probabilities are both unknown, in: Proceedings of the 7th International Conference on Knowledge Discovery and Data Mining (KDD'01), 2001, pp. 204–213.
[30] . Masashi Sugiyama, Introduction to Statistical Machine Learning 1st Edition , Elsevier
[31].http://blog.hackerearth.com/simple-tutorial-svm-parameter-tuning-python-r
[32].Data Mining: Concepts and Techniques Third Edition, Jiawei Han| Micheline Kamber|Jian Pei
[33] .https://en.wikipedia.org/wiki/Confusion_matrix
[34].

Authors:

Ch.Sarada is working as an Asst.Professor in CVR College of Engineering , Hyderabad, INDIA.
E-mail: sharada.ch@gmail.com
M.Sathya Devi is working as an Asst.Professor in CVR College of Engineering, Hyderabad, INDIA.
E-mail: satyamaranganti@gmail.com