Multi-Class Pattern Classification in Imbalanced Data

Amal S. Ghanem

Department of Computing
University of Bahrain
P.O.Box 32038, Kingdom of Bahrain
amalsaleh@itc.uob.bh

Svetha Venkatesh, Geoff West

Department of Computing
Curtin University of Technology
GPO Box U1987, Perth, 6845, Western Australia
(s.venkatesh, g.west)@curtin.edu.au

Abstract—**The majority of multi-class pattern classification techniques are proposed for learning from balanced datasets. However, in several real-world domains, the datasets have imbalanced data distribution, where some classes of data may have few training examples compared for other classes. In this paper we present our research in learning from imbalanced multi-class data and propose a new approach, named Multi-IM, to deal with this problem. Multi-IM derives its fundamentals from the probabilistic relational technique (PRMs-IM), designed for learning from imbalanced relational data for the two-class problem. Multi-IM extends PRMs-IM to a generalized framework for multi-class imbalanced learning for both relational and non-relational domains.**

Keywords-**multi-class classification; imbalanced class problem; ensemble learning;**

I. INTRODUCTION

A rich literature of pattern recognition is devoted to the techniques of multi-class pattern classification, as it is often of interest in many domains to classify more than two pattern classes. Many of these techniques are based on decomposing the multi-class problem into a set of two-class classification problems [1], [2], [3].

Despite the success of these techniques reported in different domains for various types of applications, such as text document classification, and speech recognition, most of these techniques are mainly proposed for learning from relatively balanced training data. However, in many application, the training data can be often imbalanced, where some classes of data have a small number of samples compared to the other classes, and in which it is important to accurately classify the minority cases.

The imbalanced data distribution is common in real-world problems and has resulted in serious deterioration of the performance of most well-known classification techniques [4], as a result of being biased twords the majority class and hence misclassifying most of the minority samples to be of the majority class. This imbalanced situation is even more complicated in multi-class classification, as more attention is required to handle the imbalanced situation between multiple pattern classes.

In this paper, we present our research on multi-class pattern classification in imbalanced data and present an approach, named Multi-IM to handle this problem. Multi-IM derives its basis from the relational technique PRMs-IM [5] proposed to classify two-class problems in imbalanced relational datasets by building an ensemble trained on balanced subsets.

Multi-IM extends PRMs-IM to handle the multi-class classification by employing the balancing concept of PRMs-IM in the multi-class technique All-and-One (A&O) approach [3]. The A&O approach has been proposed as a combination of the two popular multi-class approaches: One-Against-All (OAA) [1] and One-Against-One (OAO) [2], to achieve better classification results. Therefore, Multi-IM utilizes the concepts of PRMs-IM and A&O to handle the imbalanced and multi-class classification problems, respectively.

Furthermore, in our approach, we extend PRMs-IM to a generalized framework to handle the imbalanced problem in relational and flat (non-relational) domains for multi-class pattern classification.

In this paper, we evaluate our approach on a number of highly imbalanced datasets obtained from the UCI machine learning repository and a student database from Curtin University. Our experimental results show that the proposed approach achieves high performance rates in learning from imbalanced multi-class problems, and importantly, also for the minority classes.

This paper is organized as follows: In Section II, the related work is reviewed. Then, the methodology of our approach is presented in Section III, followed by the experimental results in Section. IV. Finally, Section V concludes the paper.

II. RELATED WORK

A. Learning in Imbalanced Domains

PRMs-IM [5] has been recently introduced to handle the imbalanced problem in relational datasets for the two-class problem. The main idea behind PRMs-IM is in extending a relational learning technique named Probabilistic Relational Models (PRMs) [6], to deal with the imbalanced situation. PRMs are proposed as an extension of Bayesian Networks to handle relational learning and inference by utilizing the domain relational structure in learning the probabilistic distribution. To handle the imbalanced class problem in

relational domains, PRMs-IM has been proposed as an ensemble of PRM models, in which each model is trained on a balanced subset of the training data. Each data subset is constructed from the original imbalanced dataset to include all the samples from the minority class and an equal number of samples selected randomly from the majority class. A PRM model is then learned from each subset. Once the learning phase is complete, the PRM models are combined using the weighting voting strategy, where each model may have a different weight for classifying new instances.

B. Multi-class Pattern classification

Multi-class pattern classification can be defined as finding a function F that correctly maps the input space to an output of more than two classes. Most methods designed to solve this problem are based on splitting the K-class classification problem into a number of smaller two-class subproblems. For each subproblem, an independent binary classifier is built. Then, the results of the binary classifiers are combined to get the classification result. Several techniques were proposed for decomposing the multi-class problem, including the two popular approaches: One-Against-All (OAA) [1] and One-Against-One (OAO) [2].

In the OAA approach, K binary classifiers are constructed, in which a classifier is constructed for each class. Thus, a classifier f_i is trained using the samples of class C_i against all the samples of the other classes. The results of the binary classifiers can be combined using a decision function: $F(x) = argmax_{i=1,....,K} f_i(x)$, which assigns the test sample to the class with the highest output value.

In the OAO approach, an independent binary classifier is built for each pair of classes. Thus, a classifier f_{ij} is trained using the samples of classes i and j, and hence this classifier is trained to discriminate between these two classes only. The simplest approach to combine the results of the OAO binary classifiers is majority voting, in which the test sample is assigned to the class with the highest number of votes.

By combining the two approaches, the All-and-One (A&O) approach [3] has been proposed to combine the strengths of the OAO and OAA methods and avoid the problems of each. This approach is based on the following concepts of OAO and OAA approaches: (1) for a high proportion of the miss-classifications committed by OAA, the second best output is actually the correct result, (2) the binary classifiers of OAO are highly accurate on their own, but usually lead to incorrect results when combined. Therefore, the A&O approach utilizes these observations and trains both OAA and OAO. However, for classifying new instances, the A&O approach first classifies the test sample using the OAA approach to get the first and second output classes(C_i, C_j) and then uses the corresponding OAO binary classifier f_{ij} to determine the final output.

To handle the imbalanced situation in multi-class problems, One-Against-Higher-Order (OAHO) [7] has been pro-

posed by building a hierarchy of classifiers based on the data distribution. OAHO constructs $K - 1$ classifiers for K classes in a list of $\{C_1, C_2, ..., C_K\}$. The first classifier is trained using the samples of the first class in the list C_1 against all the samples of all the other classes. Then, the second classifier is trained using the samples of the second class in the list C_2 against the samples of the higher ordered classes $\{C_3, ..., C_K\}$, and so on until the last classifier is trained for C_{K-1} against C_K. To classify new samples, a hierarchical approach is used. Thus, the sample is first classified by the first classifier. If the sample is classified as C_1, then the process terminates and the sample is assigned to class C_1. Otherwise, the second classifier is used to classify the sample, and so on till the last classifier. To resolve the imbalanced class problem in this approach, the classes are ordered in descending order based on the size of the samples in each class. This order is chosen to reduce the imbalanced situation, in which the small classes are grouped together against the majority class.

In terms of the imbalanced problem, although the popular OAA and OAO approaches have been employed successfully in different domains, their performances are significantly hindered by the imbalanced problem [8]. Similarly, the A&O approach combines the strengths of both OAA and OAO to achieve better classification results, but it has not been designed to handle the imbalanced problem. On the other hand, although OAHO has been proposed to handle the imbalanced problem for multi-class classification, its performance is sensitive by the classifier order, as miss-classifications made by the top classifiers cannot be corrected by the lower classifiers.

III. METHODOLOGY

We propose a new approach (Multi-IM) to handle the problem of learning from imbalanced multi-class domains by employing two methods: PRMs-IM and A&O. Our approach is based on extending PRMs-IM to the multi-class problem by embedding the balancing concept of PRMs-IM in A&O. Furthermore, in our approach, we aim to extend PRMs-IM to a generalized framework for relational and flat domains. Our approach is based on the idea that the balancing technique of PRMs-IM can be used for both domains, but employing the classifier that best models the given domain. Thus, PRMs can be used for relational domains and Bayesian Networks (BNs) for flat domains.

To illustrate our approach, consider a three class problem (C_1, C_2, C_3), with imbalanced data distribution. Multi-IM firstly follows the A&O approach by training OAA and OAO. For the OAA, we construct three classifiers (OAA_1, OAA_2, OAA_3), one classifier for each class. The training data of OAA_i includes all the samples of C_i as positives and all the other samples of the other classes as negatives. For the OAO, we build three classifiers $(OAO_{(1,2)}, OAO_{(1,3)}, OAO_{(2,3)})$ for each pair of classes.

Table I
THE DATA DISTRIBUTION OF THE STUDENTS DATASET.

Course	Dataset	No. Samples		
		Fail	Average	Excellent
BCom	MGT100	159	1470	86
	MKT100	88	1559	68
BCS	ST152	12	50	8
	FCS152	11	53	6
	IPE151	7	61	2

Table II
THE DATA DISTRIBUTION OF THE GLASS AND SHUTTLE DATASETS.

Dataset	No. Samples						
	C1	C2	C3	C4	C5	C6	C7
Glass	70	76	17	13	9	29	-
Shuttle	34,108	37	132	6748	2548	6	11
Shuttle-Test	11,478	13	39	2155	809	4	2

The training data $S_{(i,j)}$ of $OAO_{(i,j)}$, includes the samples of C_i and C_j as positives and negatives, respectively.

To address the imbalanced problem, the balancing concept of PRMs-IM is used in building the classifiers of the OAO and OAA. Thus, the training data D_i of each classifier f_i is used to obtain balanced subsets that include all the minority samples and a similar number of random samples of the majority class. Then, an independent classifier is trained on each balanced subset. The classifier is selected based on the domain problem, for example, using PRMs for relational domains and BNs for flat domains. These classifiers are then combined using the weighted voting strategy as applied in PRMs-IM [5] to get the result for the parent classifier f_i.

For classifying new samples, the OAA system is used to find the top two candidates (C_i, C_j). Then, the corresponding binary OAO classifier $OAO_{(i,j)}$ is used to find the final answer.

As in the A&O approach, the main issue in our approach is the large number of classifiers required, which includes $K(K-1)/2$ classifiers for OAA, K classifiers for OAA, and the classifiers of the balanced subsets that depend on the statistical distribution of the training data. However, this number of classifiers can be reduced as suggested in A&O [3] by initially training only the OAA classifiers, and after obtaining the two candidates (C_i, C_j) of OAA, the corresponding OAO classifier $OAO(C_i, C_j)$ is created.

IV. EXPERIMENTS

A. Datasets

We use the same relational student dataset used in PRMs-IM [5], which holds the data for students enrolled in the Bachelor of Computer Science (BCS) and Bachelor of Commerce (BCom) to predict student performance in second semester units given first semester results. In this paper, the attribute 'Status' that indicates a student's performance in semester II is used as a multi-class attribute. Table I shows the distribution of the training datasets for students enrolled in the period 1999-2005. In addition, the data of students enrolled in 2006 is used as an independent testing set.

In addition to the relational student dataset, we use the non-relational Glass and Shuttle datasets obtained from the UCI machine learning database. We have chosen these

datasets because they represent highly imbalanced datasets with different numbers of pattern classes. The distribution of classes in these datasets is shown in Table II.

For each dataset, 5-cross validation is performed. In addition, the algorithm was also evaluated on the separate testing sets of Students and Shuttle datasets. In these experiments, we use PRMs as the classifier for the relational Students dataset, and use the Naïve Bayes classifier for the Glass and Shuttle datasets.

B. Results

The results of Multi-IM are presented in comparison to the OAA, OAO, A&O and OAHO approaches. The Receiver Operating Characteristics (ROC) curve and the Area Under Curve (AUC) [9] are usually used to measure the performance of imbalanced classification algorithms. However, for multi-class algorithms, we need to use a multi-class AUC method. Therefore, the results are shown in terms of total AUC [10] value. In this approach, a separate AUC for each class is calculated, such that the AUC of class C_i is calculated by considering all the samples of C_i as positives and the samples of all other classes as negatives. Then, the total AUC is calculated as the summation of the AUCs weighted by the class prior probability, i.e. $AUC_{total} = \sum_{c_i \in C} AUC(c_i) * p(c_i)$, where $AUC(c_i)$ is the AUC of C_i, and $p(c_i)$ is the prior probability of C_i.

Table III shows the total AUC results obtained for the datasets, where the best result for each dataset is shown in bold. The results show that the special imbalanced algorithm (OAHO) did not perform better than the OAO and OAA algorithms, due to the problem of propagating the miss-classifications to the lower levels of the hierarchy. In addition, among the thirteen experiments shown in Table III, the A&O algorithm has generally outperformed the OAO and OAA algorithms in twelve experiments, but performed slightly worse than OAA in one experiment. On the other hand, the proposed approach (Multi-IM) has generally outperformed the other methods, including the special imbalanced algorithm (OAHO). The exception was for the Shuttle dataset, using cross validation, where Multi-IM scored less than the best score.

Table III
SUMMARY OF THE TOTAL AUC RESULTS. *: FOR THE RESULTS
OBTAINED FROM THE 5-CROSS VALIDATION, AND **: FOR THE
RESULTS OBTAINED FROM THE TESTING SETS.

Dataset	Classification Algorithm				
	OAA	OAO	OAHO	A&O	Multi-IM
Glass	0.807	0.779	0.801	0.844	**0.860**
Shuttle*	0.974	0.879	0.959	**0.997**	0.965
Shuttle**	0.947	0.941	0.985	0.981	**0.993**
MGT100*	0.733	0.718	0.727	0.740	**0.898**
MGT**	0.755	0.707	0.750	0.772	**0.895**
MKT*	0.770	0.778	0.756	0.770	**0.786**
MKT**	0.780	0.781	0.766	0.781	**0.789**
ST152*	0.793	0.757	0.786	0.801	**0.843**
ST152**	0.786	0.757	0.790	0.802	**0.845**
FCS152*	0.733	0.665	0.725	0.766	**0.904**
FCS152**	0.733	0.665	0.725	0.766	**0.904**
IPE151*	0.805	0.749	0.774	0.805	**0.883**
IPE151**	0.810	0.729	0.786	0.811	**0.897**

V. CONCLUSION

In this paper we focused on two main challenges in pattern recognition: the imbalanced class problem and multi-class classification. We reviewed the different strategies proposed to solve these two challenges. Based on this research, we outlined a framework that can handle these challenges simultaneously. Our approach (Multi-IM) is based on a relational technique designed for the binary imbalanced problem (PRMs-IM). Multi-IM extends PRMs-IM to a generalized framework for multi-class classification. The proposed approach was applied to a number of highly imbalanced datasets from different domains. The results of Multi-IM were generally better than other promising strategies and was able to predict all the classes. In this research, we focused on static datasets, our future work will involve testing this approach for temporal datasets.

REFERENCES

[1] R. Anand, K. Mehrotra, C. Mohan, and S. Ranka, "Efficient classification for multiclass problems using modular neural networks," *Neural Networks, IEEE Transactions on*, vol. 6, no. 1, pp. 117–124, Jan 1995.

[2] T. Hastie and R. Tibshirani, "Classification by pairwise coupling," *The Annals of Statistics*, vol. 26, no. 2, pp. 451–471, 1998.

[3] N. Garcia-Pedrajas and D. Ortiz-Boyer, "Improving Multi-class Pattern Recognition by the Combination of Two Strategies," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 6, pp. 1001–1006, 2006.

[4] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," *Intell. Data Anal.*, vol. 6, no. 5, pp. 429–449, 2002.

[5] A. S. Ghanem, S. Venkatesh, and G. West, "Learning in Imbalanced Relational Data," in *Proc. ICPR*. IEEE Computer Society, December 2008.

[6] N. Friedman, L. Getoor, D. Koller, and A. Pfeffer, "Learning Probabilistic Relational Models," in *Proc. IJCAI*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999, pp. 1300–1309.

[7] Y. Murphey, H. Wang, G. Ou, and L. Feldkamp, "OAHO: an effective algorithm for multi-class learning from imbalanced data," in *International Joint Conference on Neural Networks (IJCNN)*, Aug. 2007, pp. 406–411.

[8] G. Ou and Y. L. Murphey, "Multi-class pattern classification using neural networks," *Pattern Recognition*, vol. 40, no. 1, pp. 4–18, 2007.

[9] T. Fawcett, "An introduction to roc analysis," *Pattern Recogn. Lett*, vol. 27, no. 8, pp. 861–874, 2006.

[10] F. Provost and P. Domingos, "Well-trained PETs: Improving probability estimation trees. CDER Working Paper 2000-04-1S, Stern School of Business, New York University, NYU, NY 10012," 2000.