

A Modified Back-Propagation Algorithm to Deal with Severe Two-Class Imbalance Problems on Neural Networks^{*}

R. Alejo¹, P. Toribio¹, R.M. Valdovinos², and J.H. Pacheco-Sanchez³

¹ Tecnológico de Estudios Superiores de Jocotitlán
Carretera Toluca-Atlaconulco KM. 44.8, Col. Ejido de San Juan y San Agustín, 50700
Jocotitlán, Mexico

² Centro Universitario UAEM Valle de Chalco, Universidad Autónoma del Estado de México
Hermenegildo Galena No.3, Col. Ma. Isabel, 56615 Valle de Chalco, Mexico

³ Instituto Tecnológico de Toluca
Av. Tecnológico s/n Ex-Rancho La Virgen, 52140, Metepec, Mexico

Abstract. In this paper we propose a modified back-propagation to deal with severe two-class imbalance problems. The method consists in automatically to find the over-sampling rate to train a neural network (NN), i.e., identify the appropriate number of minority samples to train the NN during the learning stage, so to reduce training time. The experimental results show that the performance proposed method is a very competitive when it is compared with conventional SMOTE, and its training time is lesser.

Keywords: two-class imbalance problems, modified back-propagation, re-sampling methods and SMOTE.

1 Introduction

Back-propagation is now the most widely used tool in the field of artificial neural networks (NN). However, despite the general success of back-propagation, several major deficiencies are still needed to be solved. The major disadvantage of back-propagation is the slow rate of convergence of net output error; this is especially a major difficulty in “imbalanced” classification problems [1,2], i.e., where the training set contains many more samples of some “dominant” classes (majority classes) than the other “subordinates” classes (minority classes).

Much work has been done to deal with the class imbalance problem on neural networks [3], for example in [1] a modified back-propagation is proposed, this consists of calculating a direction in weight-space which decreases the error for each class (majority and minority class) in the same magnitude, in order to accelerate the learning rate for two-class imbalance problems. In [4,5,2,6], the error function was modified by introducing different costs associated with making errors in different classes. Basically,

^{*} This work has been partially supported under grants of: Projects 3072/2011 from the UAEM, PROMEP/103.5/11/3796 from the Mexican SEP and SDMAIA-010 of the TESJO.

when the sum of square errors is calculated, each term is multiplied by a class dependent (regularization) factor. This compensates class imbalance [4,5,6] and accelerates the convergence of the NN [2]. However, the re-sampling methods, by adding to the minority (positive) class or removing the majority (negative) class of a given data set, has become a de facto standard to counter the curse of imbalance in various domains, [7].

The re-sampling methods are the most investigated because they are independent of the underlying classifier and can be easily implemented for any problem [8]. However, the sampling methods face two common, but important, criticisms: 1) How to automatically discover the proper amount of sampling (sampling rate)? [7], and 2) in case the research a perfect balance of the training data set (TDS), in severe class imbalance problems, under-sampling involves a loss of information which can be detrimental for the classifier performance or seems inapplicable, and over-sampling modifies the data set probability distribution, increases the training time and can be prohibit its use [9].

In this paper we propose a modified back-propagation that automatically discovers the amount of over-sampling rate for a TDS, during training stage. The method is based on back-propagation Mean Square Error (MSE) to identify the over-sampling rate. The SMOTE is used in this work to re-sampling the TDS.

The remainder of the paper is organized as follows: Section 2 presents the proposed method (modified back-propagation). Section 3 discusses the methodology used in the experiments. Experimental results and a comparison are presented in section 4. Finally, section 5 presents the conclusions and outlines future research.

2 Proposed Method: Modified Back-Propagation Algorithm

In the back-propagation algorithm, the class imbalance does possess severe problems in training stage as the learning process becomes biased towards the majority class, ignoring the minority class and leaving them poorly trained at the end of the training stage. In addition, the learning process also becomes slower as it takes a longer time to converge to expect solution [1]. Empirical studies performed with the back-propagation algorithm show that class imbalance problem is given by the contribution to MSE from minority class in relation to majority class, where the most contribution to MSE is produced by the majority class. Therefore, the training process is dominated by it.

Re-sampling methods has been successfully used to deal with the class imbalance problem [7]. These methods sample the original data set, either by over-sampling the minority class or by under-sampling the majority class, until the classes are approximately equally represented. Both strategies can be applied in any learning system since they act as a preprocessing phase, thus allowing the system to receive the training instances as if they belonged to a well-balanced data set. By using this strategy, any bias of the learning system towards the majority class due to the skewed class priors will hopefully be eliminated. However, sampling methods show important drawbacks, under-sampling involves a loss of information which can be detrimental for the classifier performance and over-sampling modifies the dataset probability distribution and increases the training time [5]. In addition, in severe class imbalance problems, despite of any computational advantages of under-sampling due to the reduced sample size, under-sampling seems inapplicable in contrast to the time demanding by over-sampling [9].

Severe class imbalanced TDS can be sub-optimally over-sampling to deal with the class imbalance problem, i.e., can be find a sub-optimal over-sampling rate to reduce the class imbalance problem, and to avoid the excessive time required by the conventional over-sampling methods on severe class imbalance context.

In this work we propose a modified back-propagation algorithm that identifies automatically the amount of over-sampling rate during training stage. This proposal is based on the idea of that exist a relationship between MSE by class and the class imbalance ratio. Empirically, it has been observed that when the back-propagation algorithm is trained with class imbalanced TDS the MSE by class shows a behavior imbalanced and when it is trained with a balanced TDS the MSE by class is balanced. So we propose imbalanced *ratio* as:

$$ratio^2 = \frac{\|E^-(U)\|^2}{\|E^+(U)\|^2} * \frac{(Q^+)^2}{(Q^-)^2}, \quad (1)$$

where $E^+(U) = \frac{1}{Q^+} \sum_{q=1}^{Q^+} \sum_{j=1}^J (t_j^q - z_j^q)^2$, and $E^-(U) = \frac{1}{Q^-} \sum_{q=1}^{Q^-} \sum_{j=1}^J (t_j^q - z_j^q)^2$. J is the number of output nodes; Q^+ , Q^- number of minority and majority exemplar vectors; $\mathbf{z}^{(q)}$ is the real output vector and $\mathbf{t}^{(q)}$ desire output vector for an input vector $\mathbf{x}^{(q)}$.

The method consists in to identify during training stage the number of samples to forward propagation on the neural network, for this, first over-sampling entirely the TDS, then during the training stage the modified back-propagation finds the number of minority class samples to forward propagation on the neural network. This reduces the number of minority class samples used for trained the neural network and as result the time demanding by over-sampling is minimized. Algorithm 1 shows in detail the proposed method.

3 Experimental Set-Up

Database Description. To evaluate the effectiveness of the proposed method, experiments were carried out on four real data sets (Cayo, Feltwell, Mammography and Satimage). For data sets having more than two classes, we chose the class with fewer examples as the minority class, and collapsed the remainder as the majority class. In addition, random under-sampling was applied to obtain severe class imbalanced TDS. For each database, a 10-fold cross-validation was applied. The datasets were divided into ten equal parts, using nine folds as training set and the remaining block as test set. A brief summary is given in the Table 1.

Measures of Classifier Performance. The most traditional metric for measuring the performance of learning systems is the accuracy which can be defined as the degree of fit (matching) between the predictions and the true classes of data. However, the use of plain accuracy to evaluate the classifiers in imbalanced domains might produce misleading conclusions, since it is strongly biased to favour the majority classes [10,6]. Shortcomings of this evaluator has motivated search for new measures. One the most widely-used techniques for the evaluation of binary classifiers in imbalanced domains

Algorithm 1. Modified back-propagation

Input: N , M (number of middle neurodes), J (number output neurodes), K (number of classes); $\{x^{(q)}\}$, $t^{k(q)}$, I number of epochs; and learning rate η .

Output: the weights $\mathbf{w} = (w_{11}, w_{21}, \dots, w_{NM})$ $\mathbf{u} = (u_{11}, u_{21}, \dots, w_{MJ})$, the total and partial MSE (E, E^+, E^-) respectively.

INIT():

- 1: Read MLP file (N, M, J, Q^+, Q^-, I and η);
- 2: Generate initial weights randomly between -0.5 and 0.5 ;
- 3: Initial *ratio* = Q^+ / Q^- ;

LEARNING():

- 4: **while** $i < I$ or $E > 0.001$ **do**
- 5: $qp = 0$;
- 6: **for** $qn = 0$ to Q^- **do**
- 7: **Forward**(x^{qn});
- 8: **Update**(x^{qn});
- 9: **if** **Random**() \leq *ratio* **then**
- 10: **Forward**(x^{qp});
- 11: **Update**(x^{qp});
- 12: $qp ++$;
- 13: **end if**
- 14: **end for**
- 15: $Q^+ = qp$;
- 16: *ratio* = $(E^- / E^+) * (Q^+ / Q^-)$;
- 17: **end while**
- FORWARD**(x^q):
- 18: **for** $m = 0$ to $m < M$ **do**
- 19: **for** $n = 0$ to $n < N$ **do**
- 20: $y_m \leftarrow y_m + x_n^q * w_{nm}$;
- 21: **end for**
- 22: $y_m = \text{net}(y_m)$;
- 23: **end for**
- 24: **for** $j = 0$ to $j < J$ **do**
- 25: **for** $m = 0$ to $m < M$ **do**
- 26: $z_j \leftarrow z_j + u_{mj} * y_m$;
- 27: **end for**
- 28: $z_j \leftarrow \text{net}(z_j)$;
- 29: **end for**
- UPDATE**(x^q):
- 30: **for** $m = 1$ to M **do**
- 31: **for** $j = 1$ to J **do**
- 32: $u_{mj}^{r+1} \leftarrow u_{mj}^r + \eta \{ (t_j^{(q)} - z_j^{(q)}) [z_j^{(q)} (1 - z_j^{(q)})] y_m^{(q)} \}$;
- 33: **end for**
- 34: **for** $n = 1$ to N **do**
- 35: $w_{nm}^{r+1} \leftarrow w_{nm}^r + \eta \{ \sum_{j=1, J} (t_j^{(q)} - z_j^{(q)}) [z_j^{(q)} (1 - z_j^{(q)})] u_{mj}^{(r)} \} x_n [y_m^{(q)} (1 - y_m^{(q)})] [x_n^{(q)}]$;
- 36: **end for**
- 37: **end for**

Table 1. Data sets summary descriptions

Data set	Samples (min. - maj.)	Attributes	Original classes	Class (min. - maj.)
Cayo	133 - 5886	4	11	(2, remainder)
Feltwell	103 - 10048	15	5	(3, remainder)
Mammography	260 - 10923	6	2	(calcifications, no calcifications)
Satimage	121 - 5809	36	6	(4, remainder)

are the Receiver Operating Characteristic curve (ROC). Furthermore, a quantitative representation of a ROC curve is the area under it, which is known as AUC [11].

On other hand, the geometric mean (g -mean) is other of the most widely accepted measure criterion on class imbalance context, and is defined as $g\text{-mean} = \sqrt{a^+ * a^-}$, where a^+ , a^- are the accuracy on the minority and majority classes respectively [12]. In this work, g -mean and AUC were applied to measure classifier performance.

Neural Network Configuration. We use a MLP trained with the back-propagation algorithm, using four hidden neurons, and learning rate of 0.1. The shutdown criterion is reached when the error is smaller than 0.001 or a maximum of 20000 epochs having been performed.

Re-sampling Methods. The simplest method to increase the size of the minority class corresponds to random over-sampling, that is, a non-heuristic method that balances the class distribution through the random replication of positive examples. Nevertheless, since this method replicates existing examples in the minority class, over-fitting is more likely to occur. In ref. [13] proposed an over-sampling technique that generates new synthetic minority samples by interpolating between several pre-existing minority examples that lie close together. This method, called SMOTE (Synthetic Minority Over-sampling TEchnique), allows to the classifier to build larger decision regions that contain nearby samples from the minority class. In this work, the SMOTE is used to sampling the TDS. It adds samples until a perfect balanced distribution was reached.

4 Results and Discussion

In order to evaluate the possibilities of the proposed method, several experiments with severe class imbalanced TDS were developed. Fig. 1 shows the number of minority class samples (axis y) used in each epoch (axis x) for modified back-propagation during the training stage. It is observed that in the early epochs the modified back-propagation use the completely balanced TDS, i.e. it used all original minority class samples and that generated by SMOTE, after this the number of minority class samples are rapidly reduced. This occurs because in the first epochs of the training stage the MSE by class is imbalanced, and later the MSE by class gets a relative balance, and then needs less samples of the minority class. An immediate consequence of this is the reduction of training time, because the number of samples is less, so the processing time to an epoch is lesser.

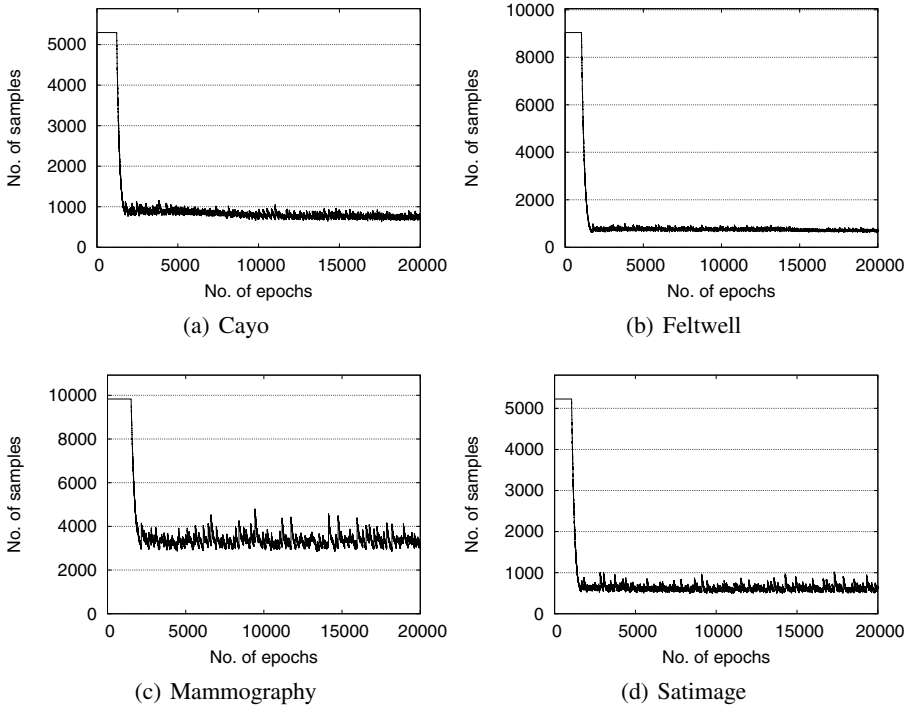


Fig. 1. Number the samples of minority class used in each epoch of training stage

Table 2 reports averaged results using AUC and g -mean for the four data sets when using the original data set to train standard back-propagation, the modified back-propagation (proposed in this work), and standard back-propagation trained with TDS sampling through of SMOTE¹. From these results, one can observe that all techniques are better than using the original imbalanced TDS and standard back-propagation. However, the SMOTE performance tends to be better than modified back-propagation.

The effects of the modified back-propagation can be better if is analysed by considering the number of samples that remain in the TDS during training stage. Results showed in Fig. 1 suggest a higher decrease in the size of the TDS when the modified back-propagation is used, whereas using SMOTE increase significantly the TDS size, because it adds information to the TDS by introducing new (non-replicate) minority classes samples, which involves a larger TDS and longer training times for the same number of training epochs.

The modified back-propagation uses on average less than 1000 minority class samples on Cayo, Feltwell, and Satimage. For Mammography needs on average less than 4000 minority class samples. SMOTE uses throughout training stage 5886, 10048, 10923, 5809 minority class samples on Cayo, Feltwell, Mammography and Satimage respectively.

¹ In the rest of the paper standard back-propagation trained with TDS sampling through of SMOTE will be referred only as “SMOTE”.

The Table 2 shows that the results obtained by modified back-propagation are not better than the obtained by SMOTE, but they are very competitive. The values of AUC and g -mean for Cayo, Feltwell, and Mammography are similar between the obtained from proposed method and the SMOTE, but the number the minority class samples used by the modified back-propagation are lesser than the used by the SMOTE. Reducing the number the samples used in the training stage, involve reach a better neural network learning time.

Table 2. Performance on four data sets measured using AUC, g -mean

Data set	AUC		
	Standard back-propagation and original TDS	Modified back-propagation	Standard back-propagation and SMOTE
Cayo	0.923(0.045)	0.967(0.029)	0.982(0.017)
Feltwell	0.934(0.042)	0.976(0.027)	0.982(0.024)
Mammography	0.764(0.035)	0.876(0.037)	0.895(0.035)
Satimage	0.659(0.081)	0.812(0.056)	0.869(0.059)
Data set	g -mean		
	Standard back-propagation and original TDS	Modified back-propagation	Standard back-propagation and SMOTE
Cayo	0.918(0.048)	0.966(0.030)	0.982(0.017)
Feltwell	0.930(0.045)	0.976(0.028)	0.982(0.024)
Mammography	0.726(0.047)	0.868(0.043)	0.892(0.039)
Satimage	0.541(0.176)	0.802(0.062)	0.863(0.065)

5 Conclusion

In this work, we propose a modified back-propagation for dealing with two-class imbalance problems. The modified back-propagation is based on the idea of that exist a relation-ship between MSE by class and the class imbalance ratio. Basically, the method consists in to identify during training stage the number the samples to forward propagation on the neural network. The modified back-propagation was compared with a conventional class imbalance technique (SMOTE). Results show that SMOTE and proposed method are very effective even with highly imbalanced data sets, but SMOTE shows a better performance. However, as we see from the results, modified back-propagation is a very competitive strategy for dealing with highly class imbalance problems, and its requirement training time are lesser.

Further research is required to investigate the potential of the strategy proposed in this paper in severe multi-class imbalance problems.

References

1. Anand, R., Mehrotra, K., Mohan, C., Ranka, S.: An improved algorithm for neural network classification of imbalanced training sets. IEEE Transactions on Neural Networks 4, 962–969 (1993)

2. Ramanan, S., Clarkson, T., Taylor, J.: Adaptive algorithm for training pram neural networks on unbalanced data sets. *Electronics Letters* 34(13), 1335–1336 (1998)
3. Zhou, Z.H., Liu, X.Y.: Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Transactions on Knowledge and Data Engineering* 18, 63–77 (2006)
4. Bruzzone, L., Serpico, S.: Classification of imbalanced remote-sensing data by neural networks. *Pattern Recognition Letters* 18, 1323–1328 (1997)
5. Lawrence, S., Burns, I., Back, A., Tsoi, A.C., Giles, C.L.: Neural Network Classification and Prior Class Probabilities. In: Orr, G., Müller, K.R., Caruana, R. (eds.) *NIPS-WS 1996*. LNCS, vol. 1524, pp. 299–314. Springer, Heidelberg (1998)
6. Oh, S.H.: Error back-propagation algorithm for classification of imbalanced data. *Neurocomputing* 74(6), 1058–1061 (2011)
7. Chawla, N.V., Cieslak, D.A., Hall, L.O., Joshi, A.: Automatically countering imbalance and its empirical relationship to cost. *Data Min. Knowl. Discov.* 17, 225–252 (2008)
8. Orriols-Puig, A., Bernadó-Mansilla, E., Goldberg, D.E., Sastry, K., Lanzi, P.L.: Facetwise analysis of xcs for problems with class imbalances. *Trans. Evol. Comp.* 13, 1093–1119 (2009)
9. Crone, S.F., Lessmann, S., Stahlbock, R.: The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing. *European Journal of Operational Research* 173(3), 781–800 (2006)
10. He, H., Garcia, E.: Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering In Knowledge and Data Engineering* 21(9), 1263–1284 (2009)
11. Fawcett, T.: An introduction to roc analysis. *Pattern Recogn. Lett.* 27, 861–874 (2006)
12. Kubat, M., Matwin, S.: Addressing the curse of imbalanced training sets: one-sided selection. In: *Proc. 14th International Conference on Machine Learning*, pp. 179–186. Morgan Kaufmann (1997)
13. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* 16, 321–357 (2002)