2017 IEEE 41st Annual Computer Software and Applications Conference

Chronic Kidney Disease Prediction on Imbalanced Data by Multilayer Perceptron

Chronic Kidney Disease Prediction

Pinar Yildirim
Department of Computer Engineering
Faculty of Engineering
Okan University
Istanbul, Turkey
pinar.yildirim@okan.edu.tr

***Abstract*—Imbalanced data is an important problem for medical data analysis. Medical datasets are often not balanced in their class labels. The traditional classifiers can be seriously affected by the imbalanced class distribution in the data. This is because they aim to optimize the overall accuracy without considering the relative distribution of each class. This study searches the effect of class imbalance in training data when developing neural network classifier for medical decision making on chronic kidney disease. Neural networks are widely used in a number of applications including data mining and decision systems. Back propagation networks are a popular type of neural networks that can be trained to recognize different patterns. The importance of these networks was considered and a comparative study of some sampling algorithms was performed based on multilayer perceptron with different learning rate values for the prediction of chronic kidney disease. This study reveals that sampling algorithms can improve the performance of classification algorithms and learning rate is a crucial parameter which can significantly effect on multilayer perceptron.**

Keywords; Imbalanced data,under sampling, over sampling, resample,smote, spread sub sample, multilayer perceptron,chronic kidney disease

I. INTRODUCTION

A well balanced dataset is very important for developing a good prediction model in classification research. Medical datasets are usually not balanced in their class distributions. An imbalanced data set contains a disproportionately high number of data in one or more classes than those for a class that is of interest. Traditional classification algorithms do not take into account the imbalance of class. They give the the same attention to the majority class and the minority class. The cost in mispredicting minority classes is higher than that of the majority class for many medical applications[1,2].

Classification is an important task of data mining and knowledge discovery in databases. A range of classification modeling algorithms, such as decision tree, nearest neighbor and neural network have been developed and successfully applied to many domains. The assumptions built into most of these algorithms are: maximizing accuracy is the goal, and the classifier will operate on data drawn from the same distribution as the training data. However, this is not always the case in real world data where one class might be represented by a large number of examples, while the other is represented by only a few [3].

Most original classification algorithms pursue to minimize the error rate: the percentage of the incorrect prediction of class labels. They ignore the difference between types of misclassification errors. In particular, they implicitly, assume that all misclassification errors cost equally. In many real-world applications, the differences between different misclassification errors can be quite large. For example, in medical diagnosis of a certain cancer, if the cancer is regarded as the positive class, and non-cancer (healthy) as negative, then missing a cancer (the patient is actually positive but is classified as negative;thus it is also called "false positive") is much more serious than the false-positive error [4].

In this study, a comparison of some sampling algorithms was performed based on multilayer perceptron with different learning rate values for the prediction of chronic kidney disease. Chronic kidney disease or chronic renal failure, as it was historically a term that encompasses all degrees of decreased renal function, from damaged at risk through mild, moderate, and severe chronic kidney failure. Chronic kidney disease is a worldwide public health problem. In the United States, there is a rising incidence and prevalence of kidney failure, with poor outcomes and high cost (see Epidemiology). Chronic kidney disease is more prevalent in the elderly population. However, while younger patients with Chronic kidney disease typically experience progressive loss of kidney function, 30%% of patients over 65 years of age with Chronic kidney disease have stable disease [5].

In medical diagnosis, neural networks are often used as a powerful discriminating classifier for the prediction of diseases[6,7]. Multilayer perceptron is one of the most popular

0730-3157/17 $31.00 © 2017 IEEE
DOI 10.1109/COMPSAC.2017.84

neural network models due to its clear architecture and comparably simple algorithm. Especially, back propagation networks are a popular type of neural networks that can be trained to recognize different patterns. Most of applications use the multilayer perceptron back propagation algorithm as the learning algorithm. The learning rate is crucial parameter for the training process for neural networks. Considering the advantages of neural networks, multilayer perceptron model using back propagation algorithm was used as a classifier to predict chronic kidney disease in this study. In addition, some experiments were performed to find an optimum learning rate so that maximum error reduction can be achieved in all iterations [8].

The main aim of this study is to help identify the challenges in imbalanced data problems in medicine and highlight the effects of learning rate parameter on multilayer perceptron model using back propagation algorithm.

II. SAMPLING METHODS

The class imbalance problem is an important issue in many data mining applications [2].Standard machine learning algorithms fail to classify imbalanced data. Due to the goal of optimizing overall accuracy, these learning algorithms prefer training instances from the majority class, and therefore reducing the predictive accuracy on the minority class.

Therefore, many researchers in machine learning area focused their attention on the class imbalance problem [9]. There are different algorithms and techniques that handle the imbalanced data sets. At the data level, the objective is to re-balance the class distribution by re-sampling the data space.

Sampling is one of the important technique manage imbalanced data. The most direct ways for dealing with class imbalance is to alter the class distributions toward a more balanced distribution. These solutions include many different forms of re-sampling such as random over-sampling, random under-sampling, improved sampling approaches and combinations of these techniques. Random over-sampling is a non-heuristic method replicate examples of the minority class in order to achieve a more balanced distribution. Random under sampling is also a non-heuristic method aim to balance the data set by eliminating examples of the majority class [3,10].

A. Under sampling

Under sampling method removes examples from the majority class to make the data set balanced. This method tries to balance the distribution of class by randomly removing majority class samples. The drawback of under sampling method is that it can discard potentially useful information that could be important for classifiers [10].Under sampling methods are divided into random and informative. Random under sampling randomly eliminates examples from the majority class till the data set gets balanced. Informative under sampling method selects only the required majority class examples based on a pre-specifies selection criterion to make the data set balanced [10].

B. Over sampling

Over sampling is a sampling approach which balances the data set by replicating the examples of minority class. The advantage of this method is that there is no loss of data in under sampling technique. The disadvantage of this technique is it may lead to over fitting and can introduce an additional computational cost if the data set is already fairly large but imbalanced [10]. Like under sampling, oversampling is also divided into two types: random oversampling and informative oversampling. Random oversampling is the method which balances the class distribution by replicating the randomly chosen minority class examples. Random over sampling generates duplicated data without creating any new information and this method is the simplest approach to over sampling where members from the minority class are chosen at random; these randomly chosen members are then duplicated and added to the new training set[1].This technique duplicates instances in the minority class[2]. Informative oversampling method synthetically generates minority class examples based on a pre-specifies criterion [10].In summary, over sampling may cause longer training time of over-fitting. The alternative to over sampling is under sampling. This approach is better than over sampling in terms of time and memory complexity. In this study, following algorithms are used for sampling:

Resample: This algorithm produces a random subsample of a dataset, sampling with replacement [11,12].

SMOTE: This algorithm creates artificial data based on the feature space similarities between existing minority examples [13]. Chawla proposed Synthetic Minority Over-Sampling Technique (SMOTE) an over-sampling approach in which the minority class is over-sampled by creating synthetic examples rather than by over-sampling with replacement. The minority class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the k minority class nearest neighbors. Depending upon the amount of over-sampling required, neighbors from the k nearest neighbors are randomly chosen[4].

Spread Sub sample: This algorithm produces a random subsample with a given spread between class frequencies, sampling with replacement[12].

III. MULTILAYER PERCEPTRON

Neural networks have emerged as a result of simulation of biological nervous system, such as the brain on a computer. Neural networks are represented as a set of nodes called neurons and connections between them. The connections have weights associated with them, representing the “strength” of those connections. Nowadays neural networks can be applied to problems that do not have algorithmic solutions or problems for which algorithmic solutions are too complex to be found. In other words, the kind of problems in which inputs and outputs variables does not have a clear relationship between them, a neural network is an efficient approach in such problems. Most neural network architecture has three layers in its structure. First layer is input layer which provides an interface with the environment, second layer is hidden layer where computation is done and last layer is output layer where output is stored. Data is propagated through successive layers, with the final result available at the output layer (Fig.1). Many different types of neural networks are available and multilayer neural networks are the most popular. Multilayer neural networks popularity is due to more then one hidden layer in its structure which help sometimes in solving complex problems which a single hidden layer neural network cannot solve[6].

A node in multilayer perceptron can be modeled as an neuron which computes the weighted sum of the inputs at the presence of the bias and passes this sum through the activation function. The whole process is defined as follows:

$$v_J = \sum_{I=1}^{p} w_{ji} X_i + \theta_j \quad (1)$$

$$y_j = f_j(v_j) \quad (2)$$

Where v_j is the linear combination of inputs $X_1, X_2, ..X_p$, θj is the bias, W_{ji} is the connection weight between the input X_i and the neuron j, and $f_j(v_j)$ is the activation function of the jth neuron and y_j is the output.

The sigmoid function is a common choice of the activation function, as defined :

$$f(a) = \frac{1}{1+e^{-a}} \quad (3)$$

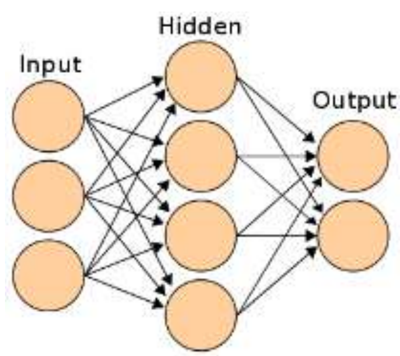

Fig. 1. Multilayer neural network

A multilayer perceptron is a feed forward neural network model that maps sets of input data onto a set of appropriate output. Back propagation is the most widely applied learning algorithm for multilayer perceptron in neural networks. Back propagation employs gradient descent to minimize the squared error between the network output values and desired values for those outputs. These error signals are used to calculate the weight updates which represent knowledge learnt in the networks. The performance of back propagation algorithm can be improved by adding a momentum term [6].

Once the architecture of multilayer perceptron has been determined, the connection weights of the network have to be computed through a training procedure based on the training patterns and the desired output. Back propagation is one of the simplest and most general methods for the supervised training of multilayer perceptron[14].

The basic back propagation algorithm works as follows[14,15]:

- Initialize all the connection weights W with small random values from a pseudorandom sequence generator.
- Repeat until convergence (either when the error E is below a preset value or until the gradient $\partial E(\mathrm{t})/\partial W$ is smaller than a preseny value).
  - Compute the update using

    $$\Delta w = -\eta \frac{\partial E(t)}{\partial w} \quad (4)$$

  - Update the weights with

    $$w(t+1) = w(t) + \Delta w(t) \quad (5)$$

  - Compute the error E(t+1)

Where t is the iteration number, W is the connection weight, and η is the learning rate. The error E can be chosen as the mean square error (MSE) function between the actual output y_j and the desired output d_j;

$$E = \frac{1}{2}\sum_{j=1}^{nj}(dj - yj)^2 \quad (6)$$

There are common training strategies: the incremental training strategy and the batch training strategy([16,17].Usually, an incremental strategy is more efficient and also faster for systems with larger training samples, as random disturbances can be induced to help the system to escape from a local minimum point.

The back propagation algorithm has some shortcomings. If the learning rate is set small enough to minimize the total error, the learning process will be slowed down. On the other hand, a larger learning rate may speed up learning process at the risk of potential oscillation. Another problem is that, partial minimal points or stable stages on error surface are often encountered during the learning process [18].

Using a momentum term is the simplest method to avoid oscillation problems during the search for the minimum values on the error surface. The weight update in back propagation algorithm with a momentum term α is defined as follows:

$$\Delta w(t) = -\eta \frac{\partial E(t)}{\partial w} + \alpha \Delta w(t-1) \quad (7)$$

where $0<\alpha<1$.

A big problem with back propagation networks is that its convergence time is usually very long. The learning rate is crucial for the training process of a two-layer neural network. Therefore, many researches have been done to find the optimal learning rate so that maximum error reduction can be achieved in all iterations [6]. If the learning rate is set small enough to minimize the total error, the learning process will be slow down. On the other hand, a larger learning rate may speed up learning process at the risk of potential oscillation. Another problem is that, partial minimal points or stable stages on error surface are often encountered during the learning process [7,18].

IV. RELATED WORK

There are different studies based on class imbalance problem on medical data and multilayer perceptron. Yan et al., developed a multilayer perceptron based decision support system to support the diagnosis of heart diseases. The input layer of the system includes 40 input variables, categorized into four groups and then encoded using the proposed coding schemes. The number of nodes in the hidden layer is determined through a cascade learning process. In the system, the missing data of a patient are handled using the substituting mean method. Furthermore, an improved back propagation algorithm is used to train the system. Their results showed that the proposed multilayer perceptron based decision support system can achieve very high diagnosis accuracy [7].

Medical datasets, as many other real-world datasets, exhibit an imbalanced class distribution. Mena et al., presented a new rule induction algorithm for machine learning in medical diagnosis and considered the poor classification accuracy caused by the classes distribution. They used a dataset for cardiovascular diseases diagnostic and three public datasets. Their experiments were performed using standard classifiers such as Naïve Bayes, C4.5 and Random Forest to compare with the new algorithm. The new algorithms overcome the other classifiers in terms of comprehensibility and validity [18-19].

Thota et al., carried experiments on classification with multilayer perceptron model using back propagation algorithm with diabetic dataset. They tried to find an optimum learning rate which is stable and takes less time for convergence [8].

Noia et al., presents a decision support systems application to predict the future occurrence of a renal failure given a set of clinical inputs. They used an ensemble of Artificial Neural Networks to predict end stage kidney disease. The model was built by training a set of neural networks via a dataset containing patients' information collected over a period of thirty-eight years (1972-2010) at Renal Unit of the University of Bari[20].

V. DATA DESCRIPTION

The data in this study was taken from UCI Machine Learning Repository which is publicly available. It includes 400 patients with 25 attributes collected from each of these patients; 250 of them have Chronic Kidney Disease(CKD). The ages of these patients vary from 2 to 90 with mean of 51:483 and a standard daviation of 17:17. The characteristics of data set is shown in Table I. As seen the Table I, the class distribution of the data set is imbalanced. It contains CKD(%62.5) and Not CKD(%37.5).

TABLE I. CHARACTERISTICS OF DATA SET

Number of Instances	400	
Number of attributes	25	
Class distribution	CKD (%62.5)	NotCKD (%37.5)

VI. EXPERIMENTAL RESULT

Chronic Kidney Disease data set was used to compare different sampling methods for the prediction of disease. Multilayer perceptron was selected to evaluate classification accuracy. Respectively, Resample, SMOTE and Spread Sub Sample algorithms were used for sampling. Same experiment was run in incremental steps by varying the learning rate between 0.1 and 0.8. WEKA 3.7.3 software was used. WEKA is a collection of machine learning algorithms for data mining tasks and is open source software. The software contains tools for data pre-processing, feature selection, classification, clustering, association rules and visualization [12].

There are many performance measures for the evaluation of the classification results, where TP/TN is the number of True Positives/Negatives instances, FP/FN is the number of False Positives/Negatives instances but some of them are used in this study. Precision is a proportion of predicted positives which are actual positive:

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

Recall is a proportion of actual positives which are predicted positive:

$$Recall = \frac{TP}{TP+FN} \tag{9}$$

Precision and recall measures are utilized to find the best method, but it is not easy to make decision. Thus, F-measure was used to get a single measure to evaluate results. The F-measure is the harmonic mean of precision and recall:

$$F-measure = \frac{2TP}{2TP+FN+FP}$$

The comparison analysis by root mean squared error was also performed and described in Table III-VI where n is the number of data patterns, $y_{p,m}$ indicates the predicted, $t_{m,m}$ is the measured value of one data point m and $\bar{t}_{m,m}$ is the mean value of all measure data points [21]. Root Mean Squared Error (RMSE) can be written as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{m=1}^{n} (y_{p,m} - t_{m,m})^2}{n}} \quad (4)$$

TABLE II. MULTILAYER PERCEPTRON WITH NO SAMPLING

Learning Rate	Precision	Recall	F-Measure	RMSE
0.1	0.998	0.998	0.998	**0.0596**
0.3	0.998	0.998	0.998	0.0622
0.5	0.995	0.995	0.995	0.0655
0.8	0.995	0.995	0.995	0.0702

TABLE III. MULTILAYER PERCEPTRON WITH RESAMPLE

Learning Rate	Precision	Recall	F-Measure	RMSE
0.1	0.998	0.998	0.998	0.0432
0.3	**0.998**	**0.998**	**0.998**	**0.0417**
0.5	0.998	0.998	0.998	0.0434
0.8	0.998	0.998	0.998	0.0451

TABLE IV. MULTILAYER PERCEPTRON WITH SMOTE

Learning Rate	Precision	Recall	F-Measure	RMSE
0.1	0.995	0.995	0.995	0.0575
0.3	0.996	0.996	0.996	0.0570
0.5	0.996	0.996	0.996	**0.0564**
0.8	0.996	0.996	0.996	0.0589

TABLE V. MULTILAYER PERCEPTRON WITH SPREAD SUB SAMPLE

Learning Rate	Precision	Recall	F-Measure	RMSE
0.1	0.997	0.997	0.997	**0.0509**
0.3	0.997	0.997	0.997	0.0514
0.5	0.997	0.997	0.997	0.0525
0.8	0.997	0.997	0.997	0.0540

Table II-V show the performance metrics of the multilayer perceptron with 10-fold cross-validation, sampling algorithms and varying the learning rate between 0.1 and 0.8. According to the table, the highest precision values were obtained for the dataset with resample sampling method with multilayer perceptron with learning rate 0.3. For example, the precision of multilayer perceptron with resample is 0.9980 which is the highest value in the Table III and has the lowest RMSE with 0.0417. These results highlighted that resample is superior to the others. Similarly, multilayer perceptron with no sampling for learning rate 0.1 and 0.3 has the same with highest precision, recall and f-measure values. As seen these tables, small learning rates can result in small RMSE values generally. However, there is no assumption that small learning rates generate small RMSE values every time.

TABLE VI. EXECUTION TIME

Learning Rate	Sampling Methods	Execution Time(Seconds)
0.1	No Sampling	7.51
	Resample	7.4
	SMOTE	10.06
	Spread Sub Sample	**5.54**
0.3	No Sampling	7.42
	Resample	7.33
	SMOTE	10.17
	Spread Sub Sample	5.68
0.5	No Sampling	7.34
	Resample	7.32
	SMOTE	10.17
	Spread Sub Sample	5.57
0.8	No Sampling	7.38
	Resample	7.37
	SMOTE	10.02
	Spread Sub Sample	5.55

The performance evaluation of multilayer perceptron with sampling algorithms was performed and the results were obtained in Table VI. This table revealed that Spread Sub Sample algorithm with learning rate 0.1 took the shortest time in the others and therefore the performance of this algorithm is better than others.

VII. DISCUSSION AND CONCLUSION

Imbalanced data set is common problem with most medical data sets. Most of machine learning algorithms can not handle imbalanced class distribution and sampling algorithms play important role in classification accuracy. In this study, a comparative experiment was carried out on sampling algorithms to predict chronic kidney disease. The effect of class imbalance in training data on performance was evaluated for multilayer perceptron. Three sampling algorithms were used to analyze the data set and their performance was evaluated by multilayer perceptron by varying learning rate. The results were evaluated on accuracy metrics and execution time. Among the sampling algorithms, Resample method has better accuracy results on the data set than the others. In terms of execution time, Spread Sub Sample algorithm took the shortest time in the others. The results highlight that sampling can significantly effect on the performance of classification studies and the learning rate parameter has to be carefully selected for any multilayer perceptron classification problem to obtain higher accuracy. In conclusion, this study makes contributions in medical data mining studies and reveals that sampling algorithms can improve the performance of multilayer perceptron with optimum learning rate parameter for learning process.

ACKNOWLEDGMENT

I would like to thank you Dr. Cinar Ceken for her help and contributions.

REFERENCES

[1] M.M.Rahman and D.N.Davis, Addressing the class imbalance problem in medical datasets, International Journal of Machine Learning and Computing, Vol.3, No.2, 2013.

[2] H.L.Yin and T.Y.Leong, A model driven approach to imbalanced data sampling in medical decision makin, Stud Health Technol Inform. 2010;160(Pt 2):856-60.

[3] Q.Gu, Z.Cai,L.Zhu&B.Huang, data mining on imbalanced data sets, International Conference on Advanced Computer Theory and Engineering, 2008.

[4] V.Ganganwa, An overview of classification algorithms for imbalanced datasets, International Journal of Emerging Technology and Advanced Engineering, vol.2,issue 4 2012.

[5] http://emedicine.medscape.com/article/238798-overview.

[6] K.Kumar and Abhishek,Artificial Neural Networks for Diagnosis of Kidney Stones Disease, I.J. Information Technology and Computer Science, 7, 20-25, 2012.

[7] H.Yan, Y. Jiang, J. Zheng,C.Peng,Q.Li "A multilayer perceptron-based medical decision support system for heart disease diagnosis" Expert Systems with Applications, vol.30, pp.272-281, 2006

[8] L.S.Thota, S.B.Changalasetty,Optimum learning rate for classification problem with MLP in data mining, International Journal of Advances in Engineering&Technology, vol.6, issue.1, pp.35-44, March 2013

[9] H.M.Nguyen, E.W.Cooper, K.Kamei ,A comparative study on sampling techniques for handling class imbalance in streaming data, SCIS-ISIS 2012.

[10] C.V.KrishnaVeni, T. R. Sobha, On the classification of imbalanced datasets, International Journal of Computer Science&Technology 2011;2:145-148

[11] P.S.Ratnoo A comparative study of instance reduction techniques. Int. Journal of Advances in Engineering Sciences 2013; 3(3).

[12] I.H.Witten, E.Frank, Data Mining: Practical Machine Learning Tool and Technique with Java Implementation. Morgan Kaufmann; 2000.

[13] H.Haibo & E.A.Garcia, Learning from imbalanced data. IEEE Transaction on knowledge and data engineering 2009;21;9: 1293-1284.

[14] R.O.Duda, P.E. Hart & D.G. Strok, Pattern classification,New York:Wiley, 2001.

[15] C.M.Bishop, Neural networks for pattern recognition, Oxford:Oxford University Press, 1995.

[16] S.Shigetoshi, F.Toshio & S.Takanori, A neural network architecture for imcremental learning, Neurocomputing, 9,111-130, 1995.

[17] J.S.R.Jang, C.T. Sun&E. Mizutani, Neuro-fuzzy and soft computing, USA, Prenctice Hall,1997.

[18] N.Baba, A new approach for finding the global minimum of error function for neural networks, Neural Networks,2,367-373,1989.

[19] L.Mena, J.A.Gonzalez Machine learning for imbalanced datasets:applications in medical diagnostic, AAAI, 2006

[20] T.D.Noia,V.C.Ostuni,F.Pesce,G.Binetti,D.Naso,F.P.Schena,E.D.Sciasco , An end stage kidney disease predictor based on an artificial neural networks ensemble, Expert Systems with Applications, vol.40,pp.4438-4445, 2013.

[21] E.U.Kuçuksille, R.Selbas,A. Şencan, Prediction of thermodynamic properties of refrigerants using data mining. *Energy conversion and management* 2011; 52: 836-848.