Decision Support Systems 48 (2010) 582–591

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

An investigation of neural network classifiers with unequal misclassification costs and group sizes

Jyhshyan Lan [a], Michael Y. Hu [b], Eddy Patuwo [b], G. Peter Zhang [c,*]

[a] College of Business Administration, Providence University, Taiwan
[b] College of Business Administration, Kent State University, Kent, Ohio 44242, United States
[c] Robinson College of Business, Georgia State University, Atlanta, GA 30303, United States

ARTICLE INFO

Article history:
Received 23 February 2009
Received in revised form 2 November 2009
Accepted 6 November 2009
Available online 13 November 2009

Keywords:
Neural networks
Group sizes
Medical diagnosis
Misclassification costs

ABSTRACT

Despite a larger number of successful applications of artificial neural networks for classification in business and other areas, published research has not considered the effects of misclassification costs and group sizes. Without the consideration of uneven misclassification costs, the classifier development will be compromised in minimizing the total misclassification errors. The use of this simplified model will not only result in poor decision capability when misclassification errors are significantly unequal, but also increase the model bias in favor of larger groups. This paper explores the issues of asymmetric misclassification costs and imbalanced group sizes through an application of neural networks to thyroid disease diagnosis. The results show that both asymmetric misclassification costs and imbalanced group sizes have significant effects on the neural network classification performance. In addition, we find that increasing the sample size and resampling are two effective approaches to counteract the problems.

1. Introduction

The classification problem arises when an investigator wishes to classify objects into one of several groups on the basis of their attribute measurements. Many business decision making situations such as financial distress detection, company performance evaluation, target marketing, production process monitoring, quality control, bond rating and credit scoring can be considered as classification problems. Classification problems also exist in many other fields such as medical diagnosis, finger print detection, and speech and hand writing recognition.

Artificial neural networks (ANNs) are one of the popular methods for classification problems [1,6–8,10]. Compared to most traditional classification approaches, ANNs are nonlinear, nonparametric, and adaptive. They can theoretically approximate any fundamental relationship with arbitrary accuracy. They are ideally suitable for problems where observations are easy to obtain but the data structure or underlying relationship is unknown. Other important features of ANNs that make them attractive for general classification problems are (1) their link to Bayes decision theory in terms of posterior probability estimation [27] and (2) their link to traditional statistical classifiers such as discriminant analysis, logistic regression, classification tree, and nearest neighbor methods [33].

Despite the growing popularity of ANNs for classification, few studies in the literature take asymmetric misclassification costs into consideration. In many cases, researchers simply make the assumption of equal misclassification costs without due process. Under this assumption, the objective of ANNs is equivalent to minimizing the total number of misclassified cases, and not the total misclassification cost. The equal cost assumption can simplify the model development and the selection of classification cutoff point. This simplification is not appropriate for situations where misclassification costs have severe unequal consequences for different groups. Depending on the situation and the perspective of the decision maker, the differences in misclassification cost can be quite large. For example, in bankruptcy prediction, a misclassification resulting from classifying a well-managed bank as an out-of-control one may have less severe consequences than misclassification derived from failure to detect an out-of-control bank for government regulators. A classification model based on equal cost assumption cannot provide enough opportunity for an early identification of potential financial decline to closely monitor those problem institutions and to take immediate corrective actions. A classification model with higher capability to detect insolvent institutions will be more appropriate for those information users, given the magnitude of the banking crisis and the enormous costs of resolution.

Medical diagnosis is another common case of asymmetric misclassification costs. Medical examiners typically assign higher cost for misclassifying a malignant tumor as a benign one to save patient's life or avoid legal issues. It is clear that in these situations, ignoring the unequal consequences of misclassification will generate bias and result in a classifier with little practical value.

* Corresponding author.
E-mail addresses: jslan@pu.edu.tw (J. Lan), mhu@kent.edu (M.Y. Hu), epatuwo@kent.edu (E. Patuwo), gpzhang@gsu.edu (G.P. Zhang).

0167-9236/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2009.11.008

Several previous studies have provided evidence that the unequal misclassification costs can significantly influence the ANN performance and optimal decision making. Kohers et al. [12,13] utilized different penalty cost functions in the context of overestimating and underestimating the actual future values to examine the effectiveness of ANNs as forecasting composite models. Salchenberger et al. [30] evaluated the ability of ANNs to predict thrift institution failures by considering the effect of different cutoff points on the Type I and Type II errors. Philipoom et al. [24] used a cost-based due-date assignment scheme to suggest that the cost of early completion may differ in form and/or degree from the cost of tardiness. They found that implicitly ignoring asymmetric consequences in the due-date assignment could be costly and ANNs could be more appropriate for problems with unequal costs for earliness and tardiness than linear programming approaches. They also suggested that ANNs can be used for a wide range of cost functions, whereas other methodologies are significantly more restricted. Berardi and Zhang [2] investigated the effect of unequal misclassification costs on neural network classification performance. Their results suggested that different cost considerations had significant effects on the neural network classification performance, particularly for smaller groups, and that appropriate use of cost information could aid in optimal decision making in a situation in which correct identification of some groups is of utmost importance.

Another issue associated with ANN application is imbalanced group sizes. The imbalanced problem occurs when there are many more instances in some groups than in others. The ability of ANNs to perform static pattern discrimination stems from their potential to create a specific nonlinear transformation into a space spanned by the outputs of the hidden units in which class separation is easier [17,18]. This transformation is constrained to maximize a feature extraction criterion, which may be viewed as nonlinear multi-dimensional generalization of Fisher's linear discriminant function. Since this criterion involves the weighted between-class covariance matrix, adaptive networks trained on a multi-group classifier problem exhibit a strong bias in favor of those classes that have the large membership in the training data. The bias toward a large group is also an undesirable feature of networks in situations where information on one particular class may be more difficult or expensive to obtain than other classes.

In practical applications, the level of imbalance can be drastic, with the ratio of the smallest group size to largest group size as high as 1 to 100, 1 to 1000, 1 to 10,000, or higher [20,25,32]. Even though it is difficult for ANNs to learn from imbalanced data sets, a large number of studies in the literature ignore the issue as though the data are balanced [4,17]. However, some previous researchers in areas such as fraud detection, telecommunications management, and oil spill detection provide evidence that the imbalanced data set can significantly influence the ANN performance and the optimal decision making [3,5,15].

Therefore, developing a neural classifier that takes into consideration both cost and group imbalance is very important for practical applications. Unfortunately, a majority of the studies in the literature focus on either cost or group imbalance and are often limited in both scope and size. Kotsiantis et al. [14] review several common methods in addressing imbalanced data sets, which include data sampling and cost-sensitive learning. Li [16] shows how a bagging ensemble variation method can be used to classify imbalanced data. Zhou and Liu [34] empirically evaluate several sampling methods in addressing training cost-sensitive neural networks. Kamimura and Uchida [9] propose a cost-sensitive greedy network algorithm with Gaussian activation functions. Peng et al. [23] use a cost-sensitive ensemble method for breast cancer diagnosis. Pendharkar [21] and Pendharkar and Nanda [22] develop neural network training methods based on threshold varying and genetic algorithm.

This research aims to explore the effects of asymmetric misclassification costs and imbalanced group sizes on ANN performance. In addition, through a comprehensive and systematic experimental study on a medical diagnosis problem (thyroid disease diagnosis), we are able to suggest strategies to deal with classification problems with significant unequal misclassification costs and uneven group distributions. In thyroid diagnosis, the goal is to determine whether a patient has a normally functioning thyroid, an under-functioning thyroid (hypothyroid), or an overactive thyroid (hyperthyroid) with a number of patients' attributes such as age, gender, and health condition, as well as results of parents' various medical tests. Thyroid diagnosis represents a difficult yet interesting classification problem because this is a three-group classification problem with extremely unbalanced group memberships. Because of the large total sample size, we are able to use a cross-validation approach to study the effect of sample size as well.

The rest of the paper is organized as follows. The next section discusses the methodology regarding research design and data sets used in this study. Results are then analyzed and reported. Finally, summary and conclusion are provided.

2. Research methodology

2.1. Data set

The data set used in this study is selected from the well-known UCI (University of California, Irvine) data repository which has been used as a benchmark for various machine learning techniques. There are 7200 cases in this thyroid disease data set, which classifies a patient as having a normally functioning thyroid, an under-functioning thyroid (hypothyroid), or an overactive thyroid (hyperthyroid). The hyperthyroid class represents 2.3% (166 cases) of the data points, the hypothyroid class accounts for 5.1% (367 cases) of the observations, while the normal group makes up the remaining 92.6% (6667 cases). The classification of thyroid level is a challenging task because the data set is highly imbalanced. For each of the 7200 cases, there are 21 attributes with 15 binary and 6 continuous variables used to determine in which of the three classes the patient belongs. These attributes represent information on patients such as age, gender, health condition, and the results of various medical tests [19,26,31]. The original data set is further divided into two parts: The test set is composed of 3700 observations while the rest of the data set is used for training purposes.

2.2. Research design

To systematically investigate the effect of unequal cost, uneven group size, and sample size on neural network classifiers, a 6 by 5 by 4 (a total of 120 cells) factorial experiment is administered. The first factor in our investigation is the misclassification cost. We consider six different levels of misclassification cost for each group. These six levels are represented by misclassification cost ratio (CR) among the three groups: (1:1:1), (3:1.5:1), (4.5:2.25:1), (7:3.5:1), (12:6:1), and (27:13.5:1), where the ratio denotes relative magnitude of one group misclassification cost over others. For example, (1:1:1) means that all misclassification costs are equal while (3:1.5:1) indicates that the misclassification cost for group 1 (hyperthyroid) is twice as severe as that for group 2 (hypothyroid) and is three times as much as that for group 3 (normal thyroid). Although many other alternative cost values can be selected based on the specific situation and consideration, these selected levels reflect a reasonable range of possible values in this study as they have significant impact on the subsequent classification decision as discussed in the next section. In addition, we choose these ratios to match the relative differences in group size levels as discussed below.

The second experimental factor is the imbalanced group size. We consider five levels of the imbalanced group size as the ratios of the sample sizes in these groups: (1:2:27), (2:4:24), (3:6:21), (4:8:18),