A Recognition-Based Alternative to Discrimination-Based Multi-layer Perceptrons

Todd Eavis and Nathalie Japkowicz
Faculty of Computer Science DalTech/Dalhousie University 6050 University Avenue Halifax, Nova Scotia Canada, B3H 1W5

Abstract. Though impressive classification accuracy is often obtained via discrimination-based learning techniques such as Multi-Layer Perceptrons (DMLP), these techniques often assume that the underlying training sets are optimally balanced (in terms of the number of positive and negative examples). Unfortunately, this is not always the case. In this paper, we look at a recognition-based approach whose accuracy in such environments is superior to that obtained via more conventional mechanisms. At the heart of the new technique is a modified auto-encoder that allows for the incorporation of a recognition component into the conventional MLP mechanism. In short, rather than being associated with an output value of "1", positive examples are fully reconstructed at the network output layer while negative examples, rather than being associated with an output value of "0", have their inverse derived at the output layer. The result is an auto-encoder able to recognize positive examples while discriminating against negative ones by virtue of the fact that negative cases generate larger reconstruction errors. A simple technique is employed to exaggerate the impact of training with these negative examples so that reconstruction errors can be more reliably established. Preliminary testing on both seismic and sonar data sets has demonstrated that the new method produces lower error rates than standard connectionist systems in imbalanced settings. Our approach thus suggests a simple and more robust alternative to commonly used classification mechanisms.

1 Introduction

Concept learning tasks represent a form of supervised learning in which the goal is to determine whether or not an instance belongs to a given class. As would be expected, the greater the number of training examples, the more reliable the results obtained during the training phase. In addition, however, we must also acknowledge that the success of supervised learning algorithms is at least partly determined by the balance of positive and negative training cases. For training purposes, then, we would consider a data set optimal if, in addition to a certain minimal size, its instances were split more or less evenly between positive and negative examples of the concept in question. This type of division would ensure that our learning algorithms are not unduly skewed in favour of one case or the other.

Unfortunately, such optimality is often hard to guarantee in practice. In many domains, it is neither possible nor feasible to obtain equal numbers of positive and negative instances. For example, the analysis of seismic data in terms of its association with either naturally occurring geological activity or man-made nuclear devices is hampered by the fact that examples of the latter are extremely uncommon (and rigidly controlled). Seismic applications are not the only ones suffering from imbalanced conditions. The problem was also documented in applications such as the detection of oil spills in satellite radar images [Kubat et al., 1998], the detection of faulty helicopter gearboxes [Japkowicz et al., 1995] and the detection of fraudulent telephone calls [Fawcett & Provost, 1997]. Thus, supervised learning algorithms used in such environments must be amenable to these inherent restrictions.

In practice, many algorithms do not perform well when the training set is imbalanced (see [Kubat et al. 1998] for an illustration of this effect). Since a significant number of real-world domains can be described in this manner, it seems logical to pursue mechanisms whose performance suffers less drastically when counter examples are relatively hard to come by. In this paper, we present preliminary results obtained via the use of a Connectionist Novelty Detection method known as auto-encoder-based classification. Essentially, auto-encoder-based classifiers learn how to recognize positive instances of a concept by identifying their common patterns. When later presented with novel instances, the auto-encoder is able to recognize cases whose characteristics are in some way similar to its positive training examples. Negative instances, on the other hand, generally have little in common with the training input and are therefore not associated with the concept under investigation.

Though the auto-encoder as just described has been successful within a number of domains, it has become clear that not all environments are equally receptive to a training phase completely devoid of counter examples. More specifically, auto-encoders tend not to be as effective when negative instances of the concept exist as a subset of the larger positive set. In such cases, the network is likely to confuse counter examples with the original training cases since it has had no opportunity to learn those patterns which can serve to delineate the two. Consequently, the method presented here will incorporate a local discrimination phase within the general recognition-based framework. The result is a network that can successfully classify mixed instances of the concept, despite having been given a decidedly imbalanced training set.

2 Previous Work

Although the imbalanced data set problem is starting to attract the attention of a number of researchers, attempts at addressing it have remained uncoordinated. Nevertheless, these research efforts can be organized into four categories

- Methods in which the class represented by a small data set gets over-sampled so as to match the size of the opposing class.