INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING, *SERIES B*
Volume 5, Number 3, Pages 238–254

USING SELF-ORGANIZING MAPS FOR BINARY CLASSIFICATION WITH HIGHLY IMBALANCED DATASETS

VINICIUS ALMENDRA AND DENIS ENĂCHESCU

Abstract. Highly imbalanced datasets occur in domains like fraud detection, fraud prediction, and clinical diagnosis of rare diseases, among others. These datasets are characterized by the existence of a prevalent class (e.g. legitimate sellers) while the other is relatively rare (e.g. fraudsters). Although small in proportion, the observations belonging to the minority class can be of a crucial importance. In this work we extend an unsupervised learning technique – Self-Organizing Maps – to use labeled data for binary classification under a constraint on the proportion of false positives. The resulting technique was applied to two highly imbalanced real datasets, achieving good results while being easier to interpret.

Key words. unsupervised learning, self-organizing maps, imbalanced datasets, supervised learning

1. Introduction

Highly imbalanced datasets arise in several real-world machine learning problems. One example is fraud detection and prediction at online auction sites. In order to keep their business growing, online auction sites like eBay need to protect buyers from unscrupulous sellers. Among the several types of fraudulent behavior that take place in online auction sites, the most frequent one is non-delivery fraud [1, 2]. The challenge faced by site operators is to identify fraudsters *before* they strike, in order to avoid losses due to unpaid taxes, insurance, badmouthing etc. In other words, for a given product listing they need to *predict* whether or not it will end up being a fraud case, in order to prevent damage. Another example is the clinical diagnosis of relatively rare but serious diseases. In this situation a false negative (someone who has the disease diagnosed as sane) is more damaging that a false positive. An important challenge when tackling these problems is the difficult interpretation of many supervised learning models. The importance of interpretable models lies in the fact that prediction in the above-stated problems is a delicate issue: usually one cannot take harsh measures against a seller just because a model predicts his listing is fraudulent, unless there is a high degree of confidence. The same holds in other domains, where false positives imply additional costs and risks. The decision to take some measure is much easier when stakeholders understand why an observation was labeled as positive.

In this paper we present an algorithm that combines an unsupervised clustering technique – the Self-Organizing Map – with the supervised learning paradigm through the use of labeled data. The proposed algorithm tackles the problem of binary classification for highly imbalanced data under a constraint on the number of false positives. Using labeled data to automatically identify clusters of observations with high probability of being positive makes the Self-Organizing Map an useful tool for exploratory data analysis, which helps understanding the data. Although we developed the proposed method focusing on fraud prediction at online auction

Received by the editors February 17, 2014.
2000 *Mathematics Subject Classification.* 62H30.

sites, it can also be used in other domains. This paper is an extended version of a previously published work [3].

In Section 2 we will present the context for our research; in Section 3 we succinctly describe Self-Organizing Maps; in Section 4 we will explain our proposed algorithm for binary classification for highly imbalanced datasets; in Section 5 we will present the experimental results, and in Section 6 we will discuss them.

2. Related Work

Two problems were highly imbalanced datasets arise are *fraud detection and prediction.* Bolton and Hand [4] did a comprehensive review regarding statistical fraud detection in several domains: credit card fraud, money laundering, telecommunications fraud, computer intrusion, and scientific fraud. Although they did not mention fraud at online auction sites, these challenges also apply. There are recently published papers specifically focused on fraud at online auction sites, some from a descriptive perspective [5–7], and others aiming fraud prediction [8–15]. Regarding the methods employed, the majority of existing works were based on supervised learning techniques: decision trees [9,10], Markov random fields [12], instance-based learners [8], logistic regression [13], online probit models [14], Adaptive Neuro-Fuzzy Inference System [16], Boosted trees [15].

Unsupervised learning methods are also used for fraud detection [4, 17]. Zaslavsky et al. [18] used Self-Organizing Maps for credit card fraud detection, creating maps to capture the typical transaction patterns and then using these maps to classify new transactions. A given transaction is classified as suspicious if the distance to its cell's weight vector is beyond some threshold. Quah et al. [19] proposed a similar system. In essence their approach is one of fraud detection through outlier identification, which does not translate directly to fraud prediction, since they are dealing with fraudulent behavior *that already happened.* The fraud prediction problem deals with identifying who might commit fraud, which is a much harder problem, since fraudsters actively try to disguise themselves as legitimate users. This is crucial in the domain of online auction sites, since fraudsters need to convince buyers (and the online auction site) of their honesty. Therefore, we do not expect a clear outlier pattern in this case.

In a previous work we adapted a technique for exploratory data analysis – Andrews curves – in order to do supervised learning [20]. In the present work we follow a similar strategy: we combine the power of Self-Organizing Maps – an unsupervised learning technique with good visualizations and thus ensuring the understandability of the results – with labeled data in order to classify new examples.

Regarding class imbalance, some common approaches to solve this problem are undersampling of the majority class, oversampling of the minority class, and SMOTE (Synthetic Minority Over-sampling Technique) [21]. Some of the abovementioned works used undersampling [8, 9], one uses an unsupervised model [12], others did not state the approach adopted [11, 13, 14].

3. Self-organizing Maps

The Self-Organizing Map [22] is a feed-forward neural network whose units are linear and topologically ordered in a bidimensional lattice of a given size. It is a model inspired in the several types of "maps" that exist in the brain of higher animals, linking for example the skin sensations of the different body portions to specific areas in the cortex [22]. Figure 3.1 depicts two examples of map topology: one with a grid topology and another with an hexagonal one.