Neurocomputing 135 (2014) 313–327

Contents lists available at ScienceDirect

Neurocomputing

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

An experimental evaluation of novelty detection methods

Xuemei Ding [a,b], Yuhua Li [a], Ammar Belatreche [a], Liam P. Maguire [a,*]

[a] School of Computing and Intelligent Systems, University of Ulster, Londonderry BT48 7JL, UK
[b] Faculty of Software, Fujian Normal University, Fuzhou 350108, China

ARTICLE INFO

Article history:
Received 20 May 2013
Received in revised form
1 September 2013
Accepted 20 December 2013
Communicated by F. Rossi
Available online 9 January 2014

Keywords:
Novelty detection
Support vector data description
Gaussian mixture
k-Means clustering
k Nearest neighbours

ABSTRACT

Novelty detection is especially important for monitoring safety-critical systems in which novel conditions rarely occur and knowledge about novelty in that system is often limited or unavailable. There are a large number of studies in the area of novelty detection, but there is a lack of a comprehensive experimental evaluation of existing novelty detection methods. This paper aims to fill this void by conducting experimental evaluation of representative novelty detection methods. It presents a state-of-the-art review of novelty detection, with a focus on methods reported in the last few years. In addition, a rigorous comparative evaluation of four widely used methods, representative of different categories of novelty detectors, is carried out using 10 benchmark datasets with different scale, dimensionality and problem complexity. The experimental results demonstrate that the *k*-NN novelty detection method exhibits competitive overall performance to the other methods in terms of the AUC metric.

1. Introduction

Novelty detection aims to identify behaviours in data that are not consistent with normal expectations [1]. It is also referred to as anomaly detection [2], outlier detection [3], concept learning [4], one-class classification (OCC) [5,6], data description [7] or single-class classification [8]. A novelty detector contains a model constructed with adequate data from the normal classes but almost none from the abnormal classes. The constructed model is then used to detect if an unseen data point is normal or novel. A novelty detector can be viewed in principle as an OCC. The difference between OCC and conventional multi-class classification is that in the former only data from one class (the normal dataset) are available.

Novelty detection is mainly based on the normal data as: (1) usually, sufficient data from normal events exist but data from abnormal events is scarce, especially in safety-critical systems where occurrence of abnormal conditions is not expected and abnormal events are difficult to model [9]; (2) even if abnormal events were available for training, they might represent only one or a few types of novelty as a novelty could be different from the past novelties and it is difficult to cover every possible abnormal event [9]; (3) the acquisition of abnormal events might be very costly [10]. So, novelty detection is often based on profiling the features which can well describe the past normal events.

There have been a number of reviews on novelty detection from differing theoretical backgrounds. An early review in 2001 discussed in Tax's Ph.D. thesis classified novelty detection into three techniques: density-based, boundary-based and reconstruction-based techniques [11]. A review conducted by Markou and Singh [12] focused on statistics-based and neural networks-based techniques for novelty detection. Another survey by Hodge and Austin [13], presented outlier detection methods from three domains, namely statistics, neural networks and machine learning. The authors classified various novelty detection methods into three main categories, namely unsupervised clustering, supervised classification and semi-supervised recognition. An overview of anomaly detection techniques, published by Patcha and Park [14], focused on novelty detection techniques in intrusion detection systems of network security. The authors reviewed a number of different novelty detection techniques including statistics-based, data mining-based and machine learning-based techniques. Another recent and comprehensive review was conducted by Chandola et al., in which the authors categorised novelty detection techniques into classification, clustering, nearest neighbours, statistical methods, information theory, and spectral theory domains [1]. Additionally, Gogoi et al. [15], provided a survey of outlier detection specifically for network anomaly identification. Then most recently, Gupta et al. [16] further provided a survey of outlier detection specifically for temporal data from the aspect of data mining.

The aforementioned reviews provide comprehensive surveys on the literature in novelty detection. These articles have contributed to the understanding of this research field and have identified some of the remaining challenges for the area. However,

* Corresponding author. Tel.: +44 28 71671165; fax: +44 28 71675470.
E-mail addresses: ding-x@email.ulster.ac.uk, xuemeid@fjnu.edu.cn (X. Ding), y.li@ulster.ac.uk (Y. Li), a.belatreche@ulster.ac.uk (A. Belatreche), lp.maguire@ulster.ac.uk (L.P. Maguire).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.12.002

there are still two issues that need to be further covered. One is that the most recent reviews focused on specific domain [15] or specific type of data [16]. There continues to be new developments and applications of novelty detection approaches. The other one is that there is also a lack of an independent and comprehensive experimental evaluation of the various different methods used by researchers working in this area. This paper aims to address these two issues by providing an updated review of recent works on novelty detection coupled with a rigorous experimental evaluation of representative novelty detection methods.

In this paper, we first provide a brief literature review of novelty detection from the perspective of the novelty detector training process, which can basically be categorised into: (1) semi-supervised training which makes use of normal data as well as artificially generated labelled or unlabelled abnormal data; and (2) unsupervised training which only utilises normal data without any labelled abnormal reference.

Following the literature review, we carefully select four representative novelty detection methods for comparative evaluation. The experimental evaluation of these methods is conducted on 10 benchmark datasets. The performance is evaluated in terms of the AUC (area under the receiver operating curve) and other statistical metrics including the average and median performance as well as the Wilcoxon signed rank test.

The remainder of this paper is organised as follows. A review of recent literature on novelty detection is presented in Section 2. Section 3 discusses the comparative evaluation experimental setup including the selection of representative novelty detectors and benchmark datasets. The experimental results and a comparative analysis are presented in Section 4. Finally, we present our conclusions in Section 5.

2. Literature review

The main objective of novelty detection is to construct a model (novelty detector), which involves a training and testing process. A brief review of the most recent related work is outlined below where different methods are classified into two main categories based on their training mode: semi-supervised training and unsupervised training techniques. While unsupervised training requires only (unlabelled) normal data, semi-supervised training of novelty detectors uses labelled or unlabelled data from both normal and abnormal classes.

2.1. Semi-supervised training

Semi-supervised training has attracted increasing interest in recent years. This technique assumes the availability of a dataset that potentially contains abnormal samples, though the label for each sample is not available a priori. It initially aims to assign labels (normal or abnormal) to each data point. The labelled data is then used to augment the training process of the novelty detector. Abnormal points are usually obtained in this approach through artificial generation of abnormal data points (e.g., [17]) or through labelling of existing abnormal points in the given dataset (e.g., [6]).

Surace and Worden [17] developed a novelty detection approach in a changing environment using the negative selection algorithm that is inspired by the human immune system. The approach basically converts a novelty detection task into a two-class classification problem. It generates points for the abnormal class based on negative selection, i.e., a randomly generated point in the input space is deemed from the abnormal class if it is not similar to any points in the normal class based on a similarity threshold. This negative point generation process continues until a sufficient number of abnormal points are obtained. Using the data for normal and abnormal classes, the novelty detection process follows a standard procedure of nearest neighbours based classification for any future unseen data points. However, the success of this approach relies on two factors: the value of the threshold used in abnormal data generation and classification, and whether the generated abnormal points completely and uniformly surround the normal class boundary. If some region of the normal class boundary is not sufficiently surrounded by the generated abnormal points, the detection of abnormal points from this region would fail.

Wu and Ye [18] used both normal and a small number of abnormal training examples to build a novelty detector to generate a small hypersphere (a closed and tight boundary) surrounding the normal data with maximised margins between the hypersphere and the abnormal data points. Their experiments validated the effectiveness of the proposed approach using the geometric mean metric. Le et al. [19] proposed an updated approach to construct an optimal hypersphere by maximising two margins at the same time, namely the inside margin between the surface of this hypersphere and the normal data and the outside margin between that surface and the abnormal data. Smola et al. [20] extended the one-class support vector machines (OCSVM) method to threshold estimates of likelihood ratios (indicating regions where novelties are more likely), and focused on finding novel instances in one set relative to the other. Both normal and created abnormal points are used by the proposed algorithm in order to solve the problem of relative novelty detection. They presented a new kernel method and their experiments showed that the proposed novelty detector outperforms OCSVM.

Instead of using SVM, Blanchard et al. [21] proposed a statistics based semi-supervised novelty detection method which has the appealing option of specifying an upper threshold on the false positive rate. An unlabelled and possibly novelty is also available at training time, i.e., they did not assume that novelties are rare and their method needs to include test samples at training time. Chen et al. [22] presented a statistical kernelised spatial depth function for novelty detection where the training set includes a mixture of normal and abnormal data with missing labels. The method estimates the depth of data points in the local structure of a data set; i.e. it detects a point as novelty if the depth of the point is below a threshold. This method has shown good detection performance in the evaluation experiment, but its training and detection speed is slow due to its high computational complexity. Catterson et al. [23] described a conditional novelty detection technique in which the model is based on Gaussian mixture and a small percentage of points in the training dataset are expected to be novelties. Khreich et al. [24] proposed an iterative Boolean combination technique for efficient fusion of the responses from multiple hidden Markov model based one-class classifiers in the receiver operating characteristics (ROC) space. Their proposed approach was tested on both synthetic and real-world host-based intrusion detection data and good performance was achieved especially for limited and imbalanced training data. Xiao et al. [25] introduced a kernel principal component analysis (KPCA) based algorithm to deal with the nonlinear relation of variables, and defined the measure of novelty. Each data point is assigned a novelty score based on which point is labelled abnormal or not. The abnormal data can be detected according to the measure of novelty. The training data used in their method are contaminated by abnormal data even though all the training data points are labelled normal.

Jiang et al. [26] defined the boundaries from a rough set based information system perspective. The dataset, which includes both normal and abnormal samples, is first presented to an information system along with the boundaries definitions. The distances between a data point and the defined boundaries are then calculated