Mechanical Systems and Signal Processing 50-51 (2015) 427–436

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Novelty detection by multivariate kernel density estimation and growing neural gas algorithm

Olga Fink [a,*], Enrico Zio [b,c], Ulrich Weidmann [a]

[a] *Institute for Transport Planning and Systems, ETH Zurich, Wolfgang-Pauli-Str. 15, 8093 Zurich, Switzerland*
[b] *Chair on Systems Science and the Energetic Challenge, European Foundation for New Energy-Electricité de France (EDF) at École Centrale Paris and SUPELEC, France*
[c] *Department of Energy, Politecnico di Milano, Italy*

ARTICLE INFO

Article history:
Received 2 May 2013
Received in revised form
25 March 2014
Accepted 28 April 2014
Available online 29 May 2014

Keywords:
Novelty detection
Multivariate kernel density estimation
Growing neural gas
Railway turnout system

ABSTRACT

One of the underlying assumptions when using data-based methods for pattern recognition in diagnostics or prognostics is that the selected data sample used to train and test the algorithm is representative of the entire dataset and covers all combinations of parameters and conditions, and resulting system states. However in practice, operating and environmental conditions may change, unexpected and previously unanticipated events may occur and corresponding new anomalous patterns develop. Therefore for practical applications, techniques are required to detect novelties in patterns and give confidence to the user on the validity of the performed diagnosis and predictions.

In this paper, the application of two types of novelty detection approaches is compared: a statistical approach based on multivariate kernel density estimation and an approach based on a type of unsupervised artificial neural network, called the growing neural gas (GNG). The comparison is performed on a case study in the field of railway turnout systems. Both approaches demonstrate their suitability for detecting novel patterns. Furthermore, GNG proves to be more flexible, especially with respect to dimensionality of the input data and suitability for online learning.

1. Introduction

Techniques of machine learning have been developed to assure that the information contained in the available dataset is used optimally and to make algorithms learn patterns efficiently, also in the case that they do not occur very frequently [17]. Main properties that are sought in the models generated by machine learning techniques are as follows: (i) generalization ability, i.e. the ability to generalize patterns from training data to previously unseen data, and (ii) fault tolerance, i.e. the ability to ignore noise in input data and assure a stable model structure with respect to small changes [10].

Cross validation [10], for example, can be applied in the training mode to determine the model structure and parameters so as to ensure that the entire dataset is exploited. Then, the robustness of the performance of the algorithms can be judged with respect to the variability in the results when applied to different subsets of the data.

Bootstrap, a resampling technique [4], can be applied to bias the density of underrepresented patterns in the dataset, so as to enable the algorithm to learn all patterns from sufficiently frequent occurrences in the training data.

* Corresponding author. Tel.: +41 44 633 27 28.
E-mail address: ofink@ethz.ch (O. Fink).

http://dx.doi.org/10.1016/j.ymssp.2014.04.022
0888-3270/

One of the underlying assumptions when using data-based methods for pattern recognition in diagnostics or prognostics is that the selected data sample used to train and test the algorithm is representative of the entire dataset and covers all combinations of parameters and conditions, and resulting system states. However in practice, operating and environmental conditions may change, unexpected and previously unanticipated events may occur and corresponding new anomalous patterns develop. Many machine learning techniques have difficulty in recognizing novel patterns that follow a different distribution or were recorded in changed operating or environmental conditions, or evolved due to some anomalous conditions that were not covered by the training dataset. Therefore, techniques are required to detect novel data patterns.

The task of detecting patterns that are different from those that the applied algorithm was trained and tested on is shared in problems such as intrusion detection [27], autonomous mobile robots detecting novelty in their environment [21] or monitoring of system's conditions and detecting anomalous conditions [25,24,11].

For the solution of these problems and others, several approaches have been proposed in the literature for novelty detection. These can be statistical approaches, in which the input data are analysed based on their statistical properties [19]. The statistical models for the analyses are subsequently used to determine if the new patterns considered are from the same distribution as the training data used to build them. Statistical approaches can be subdivided into parametric and non-parametric [19]. Parametric approaches assume an underlying distribution and determine the parameters of the distribution that best fits the data patterns. In non-parametric approaches, the form of the density function is derived from the data without a priori assuming a specific distribution [19]. Besides statistical approaches, several soft computing techniques have been applied to detect novelty in the data patterns, such as different types of neural networks [20], support vector machines [20] and artificial immune systems [3,26].

Parametric approaches have a narrow field of application, especially for high dimensional data, where the type of distribution has to be determined not only for one parameter, but for several interdependent parameters. Also some of the soft computing approaches show limitations and are for example not suitable for applications in which online updating is required. This is, for example, the case for self-organizing maps (SOMs) [20,25] for which the structure is fixed prior to the learning process.

In this paper, two types of novelty detection approaches are compared based on their application to a case study in the field of railway turnout systems: a statistical approach based on multivariate kernel density estimation and an approach based on a type of unsupervised artificial neural network, called the growing neural gas (GNG). Contrary to SOM, GNG does not require an a priori definition of network structure, but the structure evolves during the learning process based on the presented patterns. As the structure is not fixed but is adaptable, GNG can also learn new evolving patterns in an online learning process and is able to adapt to dynamically changing operating conditions.

Multivariate statistical process control analyses, monitors and diagnoses process operating performance [18,2]. Some of the approaches applied to facilitate the processing of multidimensional process parameters are similar to those applied in this study. For example, the principal component analysis is an approach to reduce the dimensionality of the data while pertaining the relevant information in the reduced number of dimensions [16].

The remainder of the paper is organized as follows. Section 2 presents the two applied approaches and their theoretical background. Section 3 describes the case study and the applied data, which are derived from the railway turnout system. Section 4 presents the evaluation of the approaches applied on the case study. Finally, Section 5 discusses the obtained results and presents the conclusions of this research.

2. Applied approaches

2.1. Selecting novelty detection algorithms

With respect to the requirements of practical applications, the criteria used to select the algorithms for this research are applicability to multi-dimensional input data, flexibility, adaptability and the ability to learn novel patterns online as they evolve. Two different approaches are selected: a statistical approach and a soft computing approach.

As there is no information available on the type and form of the underlying distribution of the input data, a non-parametric approach is selected based on kernel density estimation. Furthermore, as the dataset is multi-dimensional, multivariate kernel density estimation is applied.

From the soft computing approaches, a growing neural gas algorithm is selected due to its flexibility in the learning process and in adapting its structure to new evolving patterns.

2.2. Multivariate kernel density estimation (MVKDE)

Kernel density estimations are a flexible approach to estimate the densities of a given data distribution on which no information is available on the type of the underlying distribution [22,23]. They are also referred to as Parzen windows or Parzen-Rosenblatt windows [19]. The approach of kernel density estimation has some similarities to histogram building. One of the main differences of the construction principles of the kernel density function to those of a histogram is that the density calculation is based on an interval placed around the observed value x and not on an interval containing x that is placed around a predefined bin center [9].