Neurocomputing 126 (2014) 118–131

Contents lists available at ScienceDirect

Neurocomputing

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

Dynamic supervised classification method for online monitoring in non-stationary environments

Laurent Hartert [a], Moamar Sayed-Mouchaweh [b,*]

[a] University of Reims, CReSTIC, Moulin de la Housse, BP 1039, 51687 Reims Cedex, France
[b] University of Lille 1, Ecole des Mines de Douai, 764 boulevard Lahure, 59508 Douai, France

ARTICLE INFO

Article history:
Received 3 February 2012
Received in revised form
7 November 2012
Accepted 30 January 2013
Available online 14 June 2013

Keywords:
Dynamic pattern recognition
Classification
Incremental methods
Non-stationary processes

ABSTRACT

The monitoring of a system functioning is achieved using a classifier which determines at each instant the class of a new incoming pattern. In non-stationary environments, the classifier must be able to adjust its parameters according to changes in the environment conditions. This requires a continuous learning while new patterns are available. Incremental learning is an efficient continuous learning technique for updating the classifier parameters without starting from scratch every time a new pattern is available. However in non-stationary environments, data characteristics may drift over time. This leads to deteriorate dramatically the performance of incremental learning algorithms over time. This is due to the use of data which is no more consistent with the characteristics of new incoming data. Thus, a mechanism to use only the recent and representative patterns to update the classifier parameters without a "catastrophic forgetting" is necessary. In this paper, we propose a dynamic pattern recognition method, named Dynamic Fuzzy Pattern Matching, to be used for the online monitoring of non-stationary processes functioning. This method is based on the use of an incremental algorithm allowing to follow the accumulated gradual changes of classes characteristics after the classification of each new pattern. When the accumulated gradual changes reach a suitable predefined threshold, the classifier parameters are adapted online using the recent and useful patterns.

1. Introduction

Non-stationary processes assume different functioning modes in the course of time. Several changes in system conditions, like a leak, the wear of a tool or a bad setting, can lead the system to a fault mode. In statistical Pattern Recognition (PR) [11,19], historical patterns or observations about system functioning modes are divided into groups of similar patterns, called classes. Each class is associated to a functioning mode (normal or faulty). These patterns, with their class assignments, constitute the learning set. They are represented by a set of d features, or attributes, so they can be viewed as d-dimensional features vectors, or points, in the feature space. A supervised learning method [35] uses the learning set to build a classifier that best separates the different classes in order to minimize the misclassification error. The model of each class can be represented by a membership function which determines the membership value of a pattern to a class. Then, new incoming patterns are assigned to the class for which they have the maximum membership value. The membership function can be generated using Probability Density Function (PDF) estimation based methods or heuristic based ones. In the first category, the membership function is equal to either the PDF or to the *a posteriori* probability function. The estimation of PDF can be parametric, as the Bayesian classifier [39], or non parametric, as the Parzen window [30], voting k nearest neighbor rules [8,34] or by histograms [26]. In heuristic based methods [26], the shape of the membership function and its parameters are predefined either by experts to fit the given data set [26] or by learning to construct directly the decision boundaries as the potential functions [4], neural networks [31] or support vector machines [36].

Patterns describing the system functioning can be static or dynamic. A static pattern is represented by a point in the feature space while a dynamic pattern is represented by a multidimensional trajectory. In the latter, the feature space has an added dimension which is the time [3]. Classes can also be static or dynamic. Static classes are represented by restricted areas formed by similar static patterns in the feature space. Hence, the way in which patterns occur is irrelevant to their membership values. Therefore, the classifier parameters remain unchanged over time. However, data issued from non-stationary processes are non-stationary. In this case, classes become dynamic and their characteristics change in the course of time. Thus, classes membership functions must be adapted to take into account these temporal changes.

* Corresponding author. Tel.: +33 27712120.
E-mail addresses: laurent.hartert@univ-reims.fr (L. Hartert) moamar.sayed-mouchaweh@mines-douai.fr (M. Sayed-Mouchaweh)

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.01.047

The latter could bring new information (creation, drift, fusion, splitting of classes, etc.). This new information could concern a change in operating conditions, development of a fault or simply more significant changes in the dynamic of a process. Extracting and accommodating this new information requires an adaptive classifier with a mechanism for adjusting its parameters over time. Hence, a continuous learning over long period of time with the ability to forget data becoming obsolete and useless is needed. However, it is important that the classifier updates its parameters without a "catastrophic forgetting". Therefore, a balance between continuous learning and "forgetting" is necessary to deal with non-stationary environments.

A considerable body of research has been devoted to the design of classifiers whose operating environments are supposed to be static. Many classification methods to design an effective classifier in static environments have been developed. However, everything that exists changes over time. A typical example of changing environments is the spam detection and filtering. The descriptions of the two classes "spam" and "non-spam" evolve over time due to the changes of user preferences and "spammers" techniques to trick spam classifiers. Thus, classifiers need to adjust their parameters in order to take into account the changes in their operating environments. This self-adaptation is necessary to preserve the classifier classification accuracy. Classification in non-stationary environments is a challenging topic and active research domain for current machine learning methods.

Environments' changes can be represented by two concepts. The first concept is "concept-drift". In this concept, the underlying data distribution changes gradually over time. The second concept is "concept-evolution". In this concept, a sudden change in the underlying data distribution can manifest.

In this paper, the problem of classification in non-stationary environments is viewed in a probabilistic sense. Data characteristics are represented by their probability distributions. Thus, environments changes are observed as changes in classes' conditional probability densities.

This paper proposes an approach to design a classifier capable of detecting drifts and rapidly adapting to them in order to preserve its performances. This approach is called Dynamic Fuzzy Pattern Matching (DFPM). DFPM observes gradual changes in classes' conditional probability distributions during a time window in order to detect a drift. When, these changes become serious (greater than a predefined threshold), the classifier adapt its parameters (classes membership functions) using incremental update rule. The threshold is used to distinguish between a drift and normal fluctuations of classes' characteristics. Moreover, the drift of classes' characteristics can lead to a change in the classifier structure (number of classes). Indeed, a class description may change gradually to be very similar to another class. In this case, both classes must be merged leading to decrease the number of classes. Similarly, a class can drift leading to obtain two distinct classes. In this case, the number of classes will be incremented.

The paper is structured as follows. In Section 2, several dynamic classification methods of the literature are discussed. In Section 3, Dynamic Fuzzy Pattern Matching (DFPM) is presented as a solution for classification in non-stationary environments. In Section 4, DFPM performance is evaluated according to the ones of two well-known classification methods: Incremental Support Vector Machines (ISVM) and Incremental K-Nearest Neighbors (IKNN). The last section concludes the paper.

2. Related work

The general principle of dynamic PR methods [3,7,20,25,28] is to observe the change of some statistical properties of classes, in order to decide in which state (stable, warning or action) the system is. These states correspond, respectively, to no change, gradual change and abrupt change. Thus, the classifier parameters, *i.e.* the membership functions, will be, respectively, unchanged, slightly adapted, or relearned from scratch. The misclassification rate is one of the most used statistical properties to observe a change. In this case, data are divided into batches and their true classes are known in advance; so that the misclassification error is easy to calculate. If this rate decreases significantly after receiving a batch of patterns, then the system is in the action state and the classifier parameters must be completely relearned.

Generally, three approaches are used for the classification in dynamic environments [2,3,28,15,16,18,22,24,29]. The first one [3,30] acts directly on the classifier parameters by substituting or adding some recent and representatives patterns to the learning set according to the state (stable, warning, action) in which the system is. This adaptation is based only on the most recent batch of patterns selected by one of the two following methods. The first method uses a time window, with a fixed or a variable size, which permits to reduce or limit the growing size of the database by accepting the n most recent patterns [28,37]. The size of the time window must be well-chosen to obtain a trade-off between a fast adaptation and a sufficient number of representative patterns. The second method is based on the use of a template containing a fixed number of selected patterns according to their age and usefulness. Nevertheless, quantifying the usefulness of patterns is subjective and difficult to estimate. The second approach for adapting the membership functions is based on the use of non-stationary neural networks [1,2,7,21]. In [2], a potential function based on the distance between data points is defined for the new points. The first data point potential is considered as equal to 1 and it establishes the first neuron (or rule) which is considered as the prototype (or center) of the first cluster. Then, the next new data points may possess a potential close or greater than the one of the prototype neuron. This point can reinforce or confirm the information contained in the previous ones, or if the point is more informative than the data used as prototype, a new neuron (new rule) is added. In [1,21], the neural network is based on a multiprototype Gaussian modeling of non-convex classes. The activation function of each hidden neuron determines the membership degree of an observation to one prototype of a class. With the first acquisition, the network is initialized; there is creation of the first prototype constituting the first class. The prototype is parameterized by its center and an initial covariance matrix. Then, according to the membership degree of new acquisitions, the prototype (the hidden neuron) can be adapted, eliminated or a new prototype can be created.

The third approach is based on the use of an ensemble of M classifiers to track changes in the environment [23,9]. The classification of a pattern is achieved by taking majority vote among the classifiers in the ensemble. The classifiers are retrained in a batch mode if blocks of data are available. Each classifier has a weight according to its classification accuracy. These weights are updated by evaluating the misclassification error of classifiers to keep track on whose classification accuracy is currently the most trustworthy.

Dynamic classification methods are used to solve several real problems. In [27], a set of generated rules from a learning dataset are used to realize the fault diagnosis of automatic transmissions. These rules were no more valid after six months because of changes in the diagnosis device's parameters. The application of [3] concerns the credit-scoring which aims to decide whether a new customer is a good or a bad risk according to changes in his consumption. Guedalia et al. [17] deals with the problem of classification of the quality of fruits according to the damage resulting from bad weather or other external events. In [1], the authors aim to detect and to follow up the progressive evolution of the functioning modes of a thermal regulator due to the age of its components or to other