

Accepted Manuscript

Title: Incremental Weighted One-Class Classifier for Mining Stationary Data Streams

Author: Bartosz Krawczyk Michał Woźniak

PII: S1877-7503(15)00062-9

DOI: <http://dx.doi.org/doi:10.1016/j.jocs.2015.04.024>

Reference: JOCS 367



To appear in:

Please cite this article as: Bartosz Krawczyk, Michał Woźniak, Incremental Weighted One-Class Classifier for Mining Stationary Data Streams, *Journal of Computational Science* (2015), <http://dx.doi.org/10.1016/j.jocs.2015.04.024>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Incremental Weighted One-Class Classifier for Mining Stationary Data Streams

Bartosz Krawczyk¹ and Michał Woźniak¹

Department of Systems and Computer Networks,
Wrocław University of Technology, Wrocław, Poland.
{bartosz.krawczyk,michał.woźniak}@pwr.edu.pl

Abstract

Big data analytics, especially data stream mining, is among the most popular contemporary machine learning problems. More and more often real-life tasks could generate massive and continuous amounts of data. Standard classifiers cannot cope with a large volume of the training set and/or changing nature of the environment. In this paper, we deal with a problem of continuously arriving objects, that with each time interval may contribute new, useful knowledge to the pattern classification system. One-class classification is a very useful tool for stream analysis, as it can be used for tackling outliers, noise, appearance of new classes or imbalanced data to name a few. We propose a novel version of incremental One-Class Support Vector Machine, that assigns weights to each object according to its level of significance. This allows to train more robust one-class classifiers on incremental streams. We present two schemes for estimating weights for new, incoming data and examine their usefulness on a number of benchmark datasets. We also analyze time and memory requirements of our method. Results of experimental investigations prove, that our method can achieve better one-class recognition quality than algorithms used so far.

Keywords: data stream classification, pattern classification, one-class classification, incremental learning, machine learning

1 Introduction

Contemporary computer systems store and process enormous amounts of data. Current predictions point out, that the volume of stored information will be doubling every two years. E-mails, social webs, on-line shopping etc. produce evergrowing data, that may carry valuable hidden information. Therefore, three main issues in big data analytics must be addressed (known as 3Vs: *Volume, Velocity, Variety*): how to efficiently transfer such volumes [8], how to store them and how to extract meaningful knowledge from them [14, 16]. In this work we are mainly focusing on the *Velocity*, because designing big data analytical tools must take into consideration that most of data in modern systems arrives continuously [3] as so-called data

Download English Version:

<https://daneshyari.com/en/article/6874596>

Download Persian Version:

<https://daneshyari.com/article/6874596>

[Daneshyari.com](https://daneshyari.com)