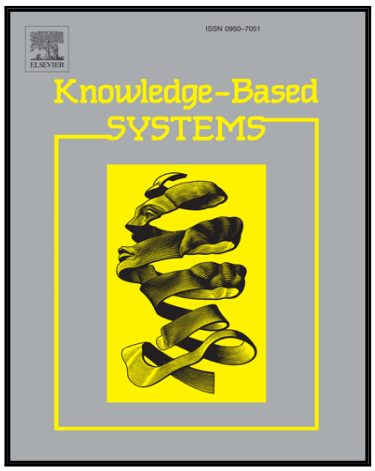

Active Learning based Support Vector Data Description Method for Robust Novelty Detection

Lili Yin, Huangang Wang, Wenhui Fan

PII: S0950-7051(18)30185-0
DOI: 10.1016/j.knosys.2018.04.020
Reference: KNOSYS 4302

To appear in: *Knowledge-Based Systems*

Received date: 19 October 2017
Revised date: 14 April 2018
Accepted date: 16 April 2018

Please cite this article as: Lili Yin, Huangang Wang, Wenhui Fan, Active Learning based Support Vector Data Description Method for Robust Novelty Detection, *Knowledge-Based Systems* (2018), doi: 10.1016/j.knosys.2018.04.020

Active Learning based Support Vector Data Description Method for Robust Novelty Detection

Lili Yin[a], Huangang Wang[a,*], Wenhui Fan[a]

[a]*Department of Automation, Tsinghua University, Beijing 100084, China*

Abstract

Practical industrial data usually has non-Gaussian data distribution and nonlinear variable correlation. Because support vector data description (SVDD) has no Gaussian limitations and can be extended to the nonlinear case by applying the kernel trick, it is one of the most widely used novelty detection methods. However, there is a great deal of actual industrial data that are mixed with much noise and uncertainty. Furthermore, SVDD may perform worse when the amount of data is too large and the data quality is poor. Describing the whole distribution with a small number of labeled samples has great practical significance and research value. This paper proposed an active learning-based SVDD method for robust novelty detection. It can reduce the amount of labeled data using an active learning framework, generalize the distribution of data and reduce the impact of noise by using the local density to guide the selection process. Experiments on two-dimensional synthetic distributions, UCI datasets and the Tennessee Eastman Process (TEP) show the effectiveness of the proposed method.

Keywords: Active Learning, SVDD, Robust Novelty Detection, TEP

1. Introduction

Novelty detection can be defined as the task of recognizing test data that differ in some respect from data available during training. The novelty detection approach is typically used when the majority of data in the training dataset is normal, but the available "abnormal" data are insufficient to construct explicit models for abnormal classes. Novelty detection can be seen as "one-class classification" (OCC) in which a model is constructed to describe the "normal" class. Novelty detection can be applied in many domains, including the fault detection in industrial system, structural damage, network intrusion detection, and disease diagnosis. This paper mainly focuses on the industrial novelty detection.

*Corresponding author

Email addresses: yll15@mails.tsinghua.edu.cn (Lili Yin), hgwang@tsinghua.edu.cn (Huangang Wang), fanwenhui@tsinghua.edu.cn (Wenhui Fan)