Arabian Journal for Science and Engineering (2022) 47:10605–10621
https://doi.org/10.1007/s13369-022-06653-4

RESEARCH ARTICLE-COMPUTER ENGINEERING AND COMPUTER SCIENCE

Disposition-Based Concept Drift Detection and Adaptation in Data Stream

Supriya Agrahari[1] · **Anil Kumar Singh**[1]

Received: 26 September 2021 / Accepted: 30 January 2022 / Published online: 8 March 2022
© King Fahd University of Petroleum & Minerals 2022

Abstract
The change in data distribution over time (known as concept drift) makes the classification process complex because of the discrepancy between current and incoming data distribution. A plethora of drift detection methods often focus on the early identification of concept drift. Along with the drift, other deformities like noise and blips are also present in the data stream. These deformities may be damaged the underlying learning system by forcing adaptation to false drift. Thereby unnecessary update performs in the learning model that leads to decrease in learner's accuracy. The existing drift detection methods are not capable of differentiating between actual and false drift. The paper proposes DBDDM, a disposition-based drift detection method, to overcome the issue of false drift. In this paper, we utilize the approximate randomization test to find the frequency of consecutive drift and compare the obtained frequency with the threshold to determine the actual drift. DBDDM compares with the several state-of-the-art methods using synthetic and real-time datasets. It exhibits a maximum increase in accuracy of 24% and 28% with a rise of 2.50 and 1.91 average ranks using Naive Bayes and the Hoeffding tree classifier, respectively.

Keywords Concept drift · Data-stream mining · Disposition based drift detection method (DBDDM) · Learning model

1 Introduction

In the growing era of technologies, a tremendous amount of streaming data generates from various applications. The streaming data may have unstable distributions [1]. The change in distribution with respect to time is known as concept drift. The data distribution change needs to be analyzed because it adds a concept drift problem in the data stream of infinite length. The concept (or context) refers to target values or classes. The data stream instances are continuous; thus, we have a short period (or single-pass) to look into the data. Sometimes, it is difficult to analyze these instances because of their characteristics like varying speed, timely ordered, and rapidly changing distribution [2].

The drift detection method is associated with the learning model [3]. The detector is used to find the significant change in the concept, whereas the learning model (or classification model) is used to forecast the outcomes of data instances (or examples). An increase in false alarm rate, decrease in accuracy of learning model, increase in classification error rate, etc., commonly identify the concept change in the data stream. Sometimes the accuracy of the classifier or predictor is degraded even the concept is stable for a long time [3]. So, it is necessary to examine such conditions in the data stream.

Concept drift detection requires in various applications such as fraud detection, cyber-security, gas sensor analysis, medical information, churn prediction, weather forecasting, etc. The applications generally use the learning model to predict the incoming data patterns. Due to concept drift, the learning model eventually becomes obsolete because it trains using old data instances. The distribution of incoming data instances changes over time. Hence, the learning model needs to be retrained using the current data distribution to minimize adverse situations like malicious activities, disasters, medical emergencies, etc.

The drift is classified as a sudden (or abrupt), gradual, incremental, recurring, blip, noise, etc., in terms of speed of change (see Fig. 1). The speed of change denotes the transition period between consecutive concepts [4]. Sudden drift occurs when an incoming data instance suddenly originates a new concept, i.e., the point of change from an old class to

✉ Supriya Agrahari
supriyaagrahari@mnnit.ac.in

Anil Kumar Singh
ak@mnnit.ac.in

1 Computer Science and Engineering Department, Motilal Nehru National Institute of Technology, Allahabad, India