



Contents lists available at ScienceDirect

Journal of King Saud University –
Computer and Information Sciencesjournal homepage: www.sciencedirect.com

Concept Drift Detection in Data Stream Mining : A literature review

Supriya Agrahari ^{a,*}, Anil Kumar Singh ^a^a Motilal Nehru National Institute of Technology Allahabad, Prayagraj, India

ARTICLE INFO

Article history:

Received 12 August 2021

Accepted 6 November 2021

Available online 3 December 2021

Keywords:

Concept drift

Concept evolution

Adaptation mechanism

Data stream mining

ABSTRACT

In recent years, the availability of time series streaming information has been growing enormously. Learning from real-time data has been receiving increasingly more attention since the last decade. Online learning encounters the change in the distribution of data while extracting considerable information from data streams. Hidden data contexts, which are not known to the learning algorithms, are known as concept drift. Classifier classifies incoming instances using past training instances of the data stream. The accuracy of the classifier deteriorates because of the concept drift. The traditional classifiers are not expected to learn the patterns in a non-stationary distribution of data. For any real-time use, the classifier needs to detect the concept drift and adapts over time. In the real-time scenario, we have to deal with semi-supervised and unsupervised data, which provide no or fewer labeled data. The motivation behind this paper is to introduce a survey identified with a broad categorization of concept drift detectors with their key points, limitations, and advantages. Eventually, the article suggests research trends, research challenges, and future work. The adaptive mechanisms are also incorporated in this survey.

© 2021 The Authors. Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Contents

1. Introduction	9524
2. Concept drift	9526
2.1. Types of drifts	9526
2.1.1. Real concept drift	9527
2.1.2. Virtual concept drift	9527
2.1.3. Class prior	9527
2.1.4. Abrupt drift	9527
2.1.5. Gradual drift	9527
2.1.6. Recurrent drift	9527
3. Detection of concept drift	9528
3.1. Concept drift detectors	9528
3.1.1. Similarity and dissimilarity based methods	9529
3.1.2. Sequential analysis based methods	9530
3.1.3. Window based methods	9530
3.1.4. Statistical based methods	9531
3.1.5. Significance analysis based methods	9532
3.1.6. Data distribution based methods	9532
3.1.7. Decision Boundary based Methods	9533
3.1.8. Model dependent based methods	9533
4. Concept drift adaptation	9535
4.1. Classification and Clustering	9535
4.2. Forgetting mechanism	9537
4.3. Windowing mechanism	9537

* Corresponding author.

E-mail address: supriyaagrahari@mnit.ac.in (S. Agrahari).