Chapter 18
Overview of Concept Drifts
Detection Methodology in Data Stream

Shabina Sayed
Jodhpur National University, India

Shoeb Ahemd Ansari
Shri Jagadish Prasad Jbnormal Tibrewala University, India

Rakesh Poonia
Bikaner Government College of Engineering, India

ABSTRACT

Real-time online applications and mobile data generate huge volume of data. There is a need to process
this data into compact data structures and extract meaningful information. A number of approaches
have been proposed in literature to overcome the issues of data stream mining. This chapter summarizes
various issues and application techniques. The chapter is a guideline for research to identify the research
issues and select the most appropriate method in order to detect and process novel class.

INTRODUCTION

Consistently, enormous volumes of tactile, value-based, and web information are ceaselessly created as
streams, which should be examined online as they arrive. The volumes of naturally produced information
are continually expanding. As indicated by (Gantz and Reinsel, 2012) more than 2.8ZB of information
were made and prepared in 2012, with an anticipated increment of 15 times by 2020. The different issues
(Mark G et. al, 1999) in handling information stream are recorded in Table 1. The incredible develop-
ment in the creation of computerized information comes about because of our encompassing condition
being furnished with an ever-increasing number of sensors. Individuals conveying advanced mobile
phones deliver information, database exchanges are being checked and put away, surges of information
are removed from virtual conditions as logs or client produced. A noteworthy piece of such informa-

DOI: 10.4018/978-1-5225-3870-7.ch018

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
Overview of Concept Drifts Detection Methodology in Data Stream

...tion is unpredictable, which implies it should be dissected progressively as it arrives. The appropriation creating the things of an information stream can change after some time. These changes, depending on the research area, are referred to as temporal evolution, covariate shift, nonstationary or concept drift (Bifet et.al, 2009). The principle qualities of the information stream display infer the accompanying limitations (Bifet, 2009):

- It is difficult to store all the information from the information stream. Just little outlines of information streams can be processed and put away, and whatever is left of the data is discarded.
- The incoming speed of information stream tuple compels every specific component to be handled essentially in real time and afterward disposed of.
- The dispersion producing the things can change after some time. Along these lines, information from the past may wind up plainly insignificant or even unsafe for the current summary.

CONCEPT DRIFT

A normal for information stream is consistent stream of information. Information measure is to a great degree expansive and possibly unending. It’s impractical to store all information. Be that as it may, significant issues identified with information stream mining are Boundless length, novel class advancement, and novel class float. Boundless length implies information stream have a vast length so require limitless length storage and training time. Concept evolution (John F. et. al, 2008) implies creating novel class and Concept drift implies information changes after some time. The most prominent case to introduce the issue of concept drift is that of distinguishing and sifting through spam email. The refinement amongst undesirable and honest to goodness messages is user specific and advances with time. Novel class does not exist on the off chance that we expect no of classes are settled. Be that as it may, sometime information stream characterization issue happen like intrusion detection, content grouping, and fault detection. So this presumption is not substantial for real streaming environment, at the point when new classes might be develop whenever. Most existing information stream characterization procedure disregards this critical part of stream information, which is the landing of a Novel Class. Concept evolution takes care of the issue of endless length and novel class. As concept drift is thought to be erratic, occasional regularity is typically not considered as a novel class issue. As a special case, if regularity is not known with assurance, it may be viewed as a concept drift issue. The center supposition, when managing the concept drift issue, is vulnerability about what’s to come. We expect that the wellspring of the objective occasion is not known with sureness. It can be accepted, evaluated, or anticipated, yet there is no assurance (ˇZliobait, 2010). (Mark G, 1999) displayed three routes in which concept drift may happen:

- Prior probabilities of classes, $P(c_1), ..., P(c_k)$ may change over time,
- Class-conditional probability distributions, $P(X|c_i), i = 1, ..., k$ might change,
Related Content

Cyber Threats in Civil Aviation
www.igi-global.com/chapter/cyber-threats-in-civil-aviation/203555?camid=4v1a

An Overview on Adaptive Group Formation Technique and the Case of the AEHS MATHEMA
www.igi-global.com/chapter/an-overview-on-adaptive-group-formation-technique-and-the-case-of-the-aehs-mathema/208796?camid=4v1a

Dependability Assessment of Two Network Supported Automotive Applications
www.igi-global.com/chapter/dependability-assessment-two-network-supported/55338?camid=4v1a

Configuring a Trusted Cloud Service Model for Smart City Exploration Using Hybrid Intelligence
www.igi-global.com/chapter/configuring-a-trusted-cloud-service-model-for-smart-city-exploration-using-hybrid-intelligence/203514?camid=4v1a