



Concept Drift in Machine Learning

Leena A Deshpande*

Department of Computer, Vishwakarma Institute of Information Technology, Pune, India

***Corresponding Author:** Leena A Deshpande, Department of Computer, Vishwakarma Institute of Information Technology, Pune, India.

Received: December 01, 2022

Published: January 10, 2023

© All rights are reserved by **Leena A Deshpande**.

Abstract

In today's era, increase in volume of data and due to variety of patterns generated, multiple challenges are raised and they need to be tackled using models based on classification theory. Many current applications face these challenges which need to be tackled with advanced solutions. Recent applications like sensor data, network traffic monitoring, stock market predictions, call centre records, web log analysers, and chemical reactor plants etc. process high amount of data where arrived data distribution may get change after certain period of time. In social media applications, users comment and share their views on social media like Twitter, Facebook causes drastic change in the behaviour and pattern as user may change their perspective and change their opinion or suddenly more hit for a particular topic or post may arrive. Such a data is referred as Data Stream. Data Stream: To handle such concept drifts in the arriving data, data stream classification is applied as a novel research problem which leads to identify change in arriving pattern. Popular algorithms of data stream mining are Classification, Clustering and Frequent pattern mining. Such novel problems opens the research challenges which addresses rigorous training of streaming data, optimum selection of algorithms, new feature selection which must be incorporated in existing machine learning algorithms.

Keywords: Data Stream; Clustering; Stock Market

Introduction

A data stream is an ordered sequence of instances that arrive continuously with time-varying intensity [1]. Due to such time varying change in pattern, a new class labels and features are also evolved. In such scenarios, normal classification based model, may fail to identify the change in the distribution and hence degrade the performance of a model. To identify such change, a degradation of performance model needs to be monitored carefully such pattern identification techniques due to change in the arriving pattern is referred as concept drift. According to Concept drift is an unforeseen substitution of one data source S_1 (with an underlying probability distribution S_1), with another source S_2 (with distribution S_2). Def. Gamma [1]. In this techniques, instead of storing the entire data instances, small summaries of data can be comput-

ed and stored. These stored instances must be processed in real time environment which remain active for particular instance of time, which may be discarded later on. Hence today's researchers are extensively paying attention for modelling the concept drift using machine learning techniques.

Bibliography

1. J Gama. "A survey on concept drift adaptation". *ACM Computing Surveys* 46.4 (2014): 1-37.