Abstract

Mining is involved with knowing the unknown characteristics from the databases or gaining of Knowledge (Knowledge Discovery) from Databases to get more useful information from the database. Real time databases which are constantly changing with time, there may arise a point when traditional Data Mining techniques may not be adequate as there may be a previously unknown class label involved or new properties of data which need to be taken into consideration. Thus as time passes and new data is in the dataset, the model predicted by the data mining techniques may become less accurate. This phenomenon is known as Concept Drift. The meaning of Concept Drift is the statistical properties of the target variable, i.e. how the properties of the target variable change over the course of time. The basic idea behind the Mining Concept Drift from Data Stream by Unsupervised Learning is to detect the Concept Drift present in the Data Stream, which is used in majority of Web-Based Applications like Fraud Detection & Spam E-mail Filtering etc. The approach taken here is both for the Offline Approach & an Online Approach, which can be easily merged with the current Web-Based Applications. Some examples for Concept Drift are – In a fraud detection application the target concept may be a binary attribute FRAUDULENT with values "yes" or "no"; or "yes"; that
indicates whether a given transaction is fraudulent. Or, in a weather prediction application, there may be several target concepts such as EMPERATURE, PRESSURE, and HUMIDITY. Each of these target parameters change over time and over model should be able to accommodate these changes or the Concept. In order to overcome the problems of the Offline or Desktop based processing to detect the Concept drift (which is available), it is aimed here to move the Concept Drift Detection process to the Cloud (web) & have it for Web-Based Applications too.

References

- Guénaël Cabanes and Younès Bennani, “Change detection in data streams through unsupervised learning?”, WCCI 2012 IEEE World Congress on Computational Intelligence, 2012.
- W. Nick Street and Yong Seog Kim. A Streaming Ensemble Algorithm (SEA) for Large-Scale Classification. KDD – 01. San Francisco, CA.
- W. Nick Street and Yong Seog Kim. A Streaming Ensemble Algorithm (SEA) for Large-Scale Classification. KDD – 01. San Francisco, CA.
- C. C. Aggarwal, J. Han, J. Wang, and P. S. Yu, “A framework for clustering evolving data streams,” in Very Large Data Base, 2003, pp. 81–92.

Index Terms
Keywords
Concept Drift  Data mining  Data Stream.