

Tutorial on Modeling Changes in Evolving Data

Under the framework of the 10th International Conference on Intelligent Systems
Design and Applications, ISDA '10
November 29 – December 1, 2010, Cairo, Egypt
Conference web page: <http://cig.iet.unipi.it/isda2010>

Tutorial Presenter
Vasudha Bhatnagar,

Department of Computer Science
University of Delhi, Delhi 110007.

Email: vbhatnagar@cs.du.ac.in,

<http://people.du.ac.in/~vbhatnagar/>

Abstract

Learning/Mining from data for descriptive or predictive purposes is the goal of intelligent data analysis. Learning in presence of concept drifts has been addressed effectively by machine learning community in last two decades. With the increasing presence of evolving datasets in scientific and commercial database application, the problem of model maintenance has been gaining importance. However, most of the work to the best of authors knowledge focuses on the problem of learnability and maintaining the currency of the learnt/discovered model as the concepts vary with time. In this tutorial generalization of this approach to descriptive models will be presented. The tutorial presents a generic framework for modeling changes in the patterns/concepts over time. Challenges and tasks involved in the process of change modeling will be discussed in detail. The process is instantiated with the change models developed for descriptive patterns viz. clustering and frequent itemsets.

Introduction

Data streams represent an important genre of real life database applications and have an inherent temporal dimension. Examples from applications (telecommunications, networks, traffic management, trend-related analysis, web-click streams, intrusion detection, and sensor networks etc.) where data arrives continuously forming an unbounded stream of data. These applications give rise to evolving data sets hiding changing patterns, which are manifestation of the underlying data generating process. The non-stationary nature of the process can be captured by continuous monitoring of the patterns and detecting the changes therein.

A pre-requisite for detection of changes is identification of the types of changes that one is looking for. This necessitates designing a change model in the context of specific characteristic(s) of the DGP. Questions like the following arise i) how the change model forms the basis for describing and quantifying changes ii) how to detect changes at multiple level of abstractions at multiple time

granularities, iii) who will be users of change model, iv) what are the different facets of change model, v) how to handle user subjectivity? etc..

It is therefore important to recognize the problem of change modeling and identify the tasks involved therein. Design issues for the functions for accomplishing each task have to be addressed appropriately. Since the change model has to be designed to meet the user requirement, a data mining practitioner/analyst may be required to design the model to suit the specific requirement.

In this tutorial a generic architecture for designing an intelligent system to detect changes in evolving data sets will be presented. The issues and challenges mentioned above will be discussed. Finally as an example, change model for streaming itemsets and streaming clusters will be presented.

Outline of the Tutorial Topics

During the tutorial (half day) following topics will be covered:

- What does evolving data and data generation process mean?
- What does change modeling mean?
- Formulation of change model for streaming itemsets
- Formulation of change model for clustering in data streams
- Change detection in classification model

Relevant references will be made available

Intended Audience

Since intelligent data analysis is the goal of machine learning and data mining community, the topic is of interest to researchers and practitioners from both areas. Particularly, those involved in learning/mining in data stream environment or other evolving datasets will find it practically useful. Those who practice business intelligence will feel motivated to develop tailored models for detecting changes in the patterns discovered over time.

Biography of the Presenter

I did my masters in Computer Applications from University of Delhi, Delhi, India in 1985. I worked in Centre for Development of Telematics from 1985 - 1989 as a software engineer in Operating System and Traffic group. Thereafter I taught UG courses in MotiLal Nehru College, UoD. I completed my doctoral studies from Jamia Millia Islamia, New Delhi, India in 2001. I am currently an Associate Professor in the Department of Computer Science, University of Delhi, Delhi, India. My broad area of interest is Intelligent Data Analysis. I am particularly interested in developing process models for Knowledge Discovery in Databases, and data mining algorithms for classification, clustering and different types of item-sets. I am currently working on problems pertaining to modeling of changes in discovered knowledge in evolving (streaming) data sets, handling user subjectivity in KDD, projected clustering, outlier detection, classification of imbalance data sets, classification and cluster ensembles. Recently, I am also looking into application of these techniques in astronomy and bio-informatics.

Contact Information

For further information, please contact Vasudha Bhatnagar (vbhatnagar@cs.du.ac.in)