Second Order Data Mining of Financial Transaction Time-Series for Fraud Detection and Improved Customer Relationship Management

Speaker:
Dr. Pawan Lingras
Professor and Director of M.Sc. in Computing and Data Analytics
Department of Mathematics and Computing Science
Saint Mary's University, Halifax, Nova Scotia, B3H3C3
email: pawan@cs.smu.ca

Abstract

Clustering is an unsupervised learning technique. Since it does not require an expert categorization of patterns, it is one of the first data mining techniques applied to a new dataset. It helps a data scientist to identify patterns in a dataset. There is no correct solution for a clustering problem. The aim of clustering is to assign objects that are alike to the same cluster and ensure that different clusters are well separated from each other. This paper proposes a novel technique for time-series data mining that uses multiple and sequential application of a number of data mining techniques to gain insights into financial transaction dataset that can be used for fraud detection and other behavioral aspects of a customer.

An important first step in creating unsupervised profiles of a customer are finding an appropriate representation. Time-series of financial transactions offer us multiple alternatives:

1. Spending distribution from month to month (Jan, Feb, …, Nov, Dec)
2. Spending distribution from day to day in a week (Monday, Tuesday, …, Saturday, Sunday)
3. Spending distribution during a day (morning, afternoon, evening, night)
4. Spending distribution among various categories (grocery, household, restaurants)
5. Geographical distribution of spending:
    - Spending in City of Residence
    - Spending in Province of Residence
    - Spending in Country of Residence
    - Online spending
    - Histogram of spending based on distance of vendor from the residence

Each one of these representations describe different behavioral aspects of a customer. To generate meaningful insights from this first order of data mining, the project proposes a second order data mining. The five set of profiles can be used as attributes in a derived dataset. The dataset can then be further analyzed using various data mining techniques including business intelligence, association mining, supervised learning (classification), and second order unsupervised learning (clustering).

Business intelligence will help us compare summary statistics of different profiles against the population. For example, "value conscious" customers spend a greater percentage of their money in groceries than average.

The association mining will provide rules such as:

- Those who spend more in summer tend to spend more on weekends
- Those who do not spend in summer tend to spend more on hotels in February

The supervised learning will use known incidences of frauds and anomalous behavior in the financial transaction data and create models that predict the chances of frauds and anomalies based on the profiles of the customers.

The second order unsupervised learning will also help us understand the correlations between different profiles. For example, "summer spenders" tend to be "non-local spenders".
The resulting insights will help us identify:

- Unusual spending pattern such as a client who never spends at night having a large transaction
- Building an inference system that uses the association rules that make up normal behavior. Any transaction that fall outside normal behavior can be easily assigned a probability of potential fraud
- Additional insights may include knowledge such as:
    - The customer only uses this card for restaurants and hotels. That means, she must be using a different card for other spending needs
    - A customer with this profile should be spending more on travel. The low or absence of travel spending on this card may suggest a use of another card

  These insights can be used to encourage customer to shift spending from other cards to this card.

About the speaker:

Pawan Lingras is a graduate of IIT Bombay with graduate studies from University of Regina. He is currently a Professor and Director of Computing and Data Analytics at Saint Marys University, Halifax. He is also internationally active having served as a visiting professor at Munich University of Applied Sciences, IIT Gandhinagar, as a research supervisor at Institut Superieur de Gestion de Tunis, as a Scholar-in-Residence, and as a Shastri Indo-Canadian scholar. He has delivered more than 35 invited talks at various institutions around the world. He has authored more than 200 research papers in various international journals and conferences. He has also co-authored three textbooks, and co-edited two books and eight volumes of research papers. His academic collaborations/co-authors include academics from Canada, Chile, China, Germany, India, Poland, Tunisia, U.K. and USA. His areas of interests include artificial intelligence, information retrieval, data mining, web intelligence, and intelligent transportation systems. He has served as the general co-chair, program co-chair, review committee chair, program committee member, and reviewer for various international conferences on artificial intelligence and data mining. He is also on editorial boards of a number of international journals. His research has been supported by Natural Science and Engineering Research Council (NSERC) of Canada for twenty-five years, as well as other funding agencies including NRC-IRAP and MITACS. He has also served on the NSERC's Computer Science peer review committee. He has been awarded an Alumni association excellence in teaching award, Student union's faculty of science teaching award, and President's award for excellence in research at Saint Mary's University.