A model to discover rare patterns using ML techniques

DBSE - TeamProject-2020
Supervisor - Sadeq Darrab
Team - 11
OVGU

Agenda

1. Motivation
2. Research Aim
3. Early literature review results
4. Tentative Timeline

Motivation

Fig 1

Fig 2

Rare Patterns & Challenges

Fig 3

Fig 4

Rare patterns appears in less percentage of data

Hard to generalise

State of the art algorithms misses rare patterns because these rules has less than min support

If we lowers the minsupport number of rules generated explodes

So far - >

Why it doesn't work?

Almost all methods suffers from performance Issues, no works for all method!

Streaming data

Noise or rare?

Scalability and performance

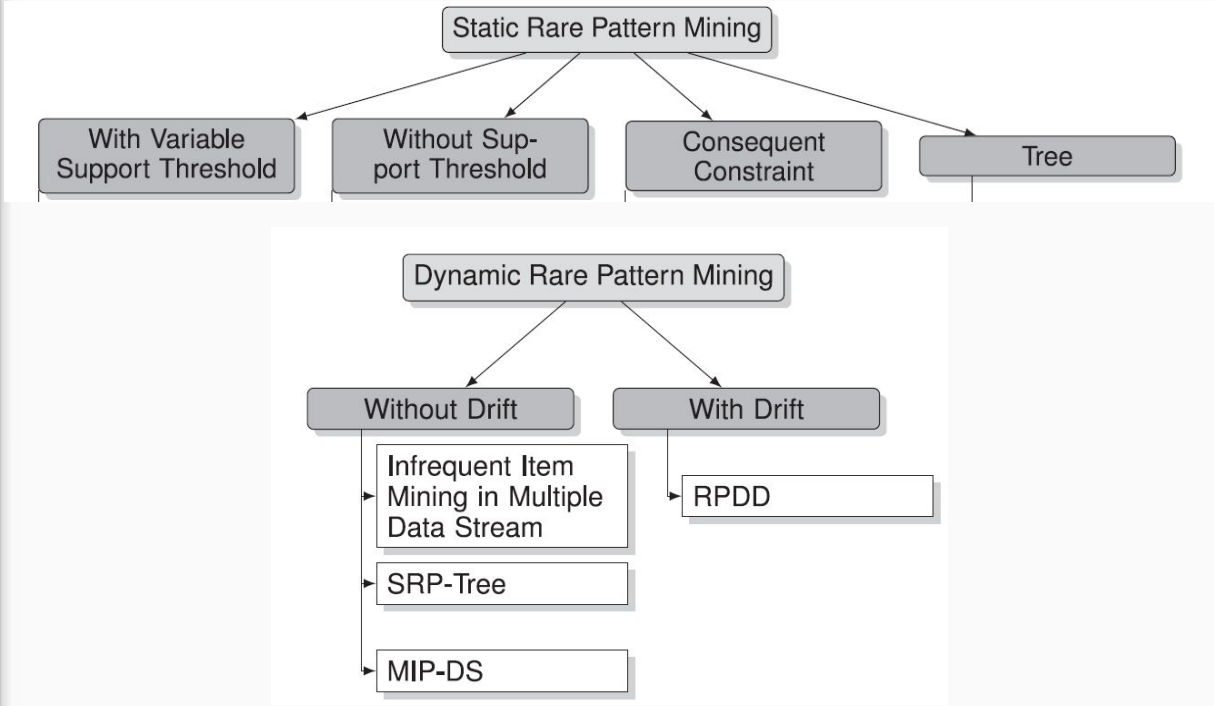

Fig 5,6

Lets use clustering techniques!

Approaches proposed-

1. Same techniques as mentioned in paper- “Clustering association rules to build beliefs and discover unexpected patterns” And use OPTICS instead of “DBSCAN” for clustering and compare
2. Use approach from paper “Generative Adversarial Active Learning for Unsupervised Outlier Detection” to find outliers.
3. Apply apriori and then clustering algorithm only on the data with high confidence and below min support and then compare results with paper “Clustering association rules to build beliefs and discover unexpected patterns”. This should detect noise and rare patterns both.

Early Literature Review

1. Rare itemsets mining algorithm based on RP-Tree and Spark framework

2. Modern Applications and Challenges for Rare Itemset Mining

3. Mining Frequent Patterns with Multiple Item Support Thresholds in Tourism Information Database

4. Mining Association Rules with Multiple Minimum Supports

5. Rare Association Rules Mining of Diabetic Complications Based on Improved Rarity Algorithm

6. Mining Rare Patterns by Using Automated Threshold Support

7. Clustering association rules to build beliefs and discover unexpected patterns

8. Unsupervised Rare Pattern Mining: A Survey

9. Pattern Mining Predictor System for Road Accidents

10. Rare pattern mining: challenges and future perspectives

11. Generative Adversarial Active Learning for Unsupervised Outlier Detection

12 . Development of a New Metric to Identify Rare Patterns in Association Analysis: The Case of Analyzing Diabetes Complications

Timeline

Team

Madhu

Deeksha

Stanley

Seles

Surabhi

Priyam

Will Exchange roles and responsibility iteratively in each sprint.

Reference-

Fig 1 to 4

www.google.com

Fig , 5,6

Yun Sing Koh and Sri Devi Ravana. 2016. Unsupervised rare pattern mining: A survey. ACM Trans. Knowl. Discov. Data 10, 4, Article 45 (May 2016), 29 pages.
DOI: http://dx.doi.org/10.1145/2898359

More to add...