Analytical Study for Pattern Mining in E-Commerce Data

Nafisa Shaikh, Nikita Medid, Shilpi Singh & Anushree

SAE, Kondhwa, Pune, India.

Abstract:

If Analytical Study for Pattern Mining of E-Commerce did not exist then one would not be able to extract interesting patterns for business boost. DDoS is Detection System, [CBR] is Case Based Reasoning, [IDS] is Intrusion Detection System are few of the systems available in the market. k-nearest neighbors Rapid Retrieval algorithms, Genetic algorithm based approach, string matching algorithm are some of the different algorithms used in the above systems. The above mentioned systems do the cover the overall domain properly, appropriate cases may not be retrieved, extensive “training sets” are often required to characterize the pattern. In our system the input data will be preprocessed then by using Info-Gain and algorithms like Aprori and Eclat it will give the best output i.e. the best interesting patterns.

Keywords
1. TF-IDF,
2. Stopword,
3. Info-Gain,
4. Frequent Item Set,
5. Support and confidence,
6. Association Rule,

1. Introduction

Our System is basically for the convenience and to create an effortless scenario for a business boost, The overall input data consisting of tedious protracted data it is extended or elongated. Then this input data will be preprocessed by removing stopwords and then by using stemming. After the preprocessed data by Shannon info-gain it will get more compact and we will get frequently occurring dataset. The methodologies used in our system are Apriori algorithm and Eclat algorithm. Apriori uses a bottom up approach, in which frequent subsets are extended by one item at a time and then groups are tested against the data, Apriori is terminated when no further successful extensions are found. It uses BFS and Hash tree structure. In Eclat Algorithm, it finds the elements by means of DFS. It is more simple and efficient as compared to Apriori algorithm. It uses database in vertical form. It cannot use horizontal database as Apriori algorithm does. If the database is in horizontal form then it converts it into vertical form. Like in apriori algorithm we don’t need to scan the given input database once more again and again. It scans the input database only once. Support is calculated and counted as done in Apriori Algorithm. But here confidence is not calculated.

2. Related Work

1. Discovering potential user browsing behavior using custom built apriori algorithm: This was used for educational log file, to find relations between them, different co relations and rules. This system has three main phases namely data preprocessing, pattern analysis and pattern discovery. This system is based on real time issues and is very promising. The system helps the user to take selective decisions based on the rules obtained from this system. The main contributions of this are:

1) Data Preprocessing: Here the raw database is preprocessed such that only
the essential data from the database is extracted for its useful use.

2) **Pattern Analysis**: After getting the the sorted data from the database the data is analyzed so as we can get or create a pattern for discovery of frequent data item sets.

3) **Pattern Discovery**: After the data is analyzed from the dataset we get a list of all the data items which have occurred frequently depending upon various frequency and in this way different patterns are discovered.

### 2. An apriori algorithm in distributed data mining system:
Here the distribution system is an optimized distributed system. In this system communication cost is reduced and decreased by using distributed information in distributed surroundings. In this system the response time is improved by measuring the response time with the communication and computation factors. Therefore here the data is mined by using apriori algorithm in distributed system.

### 3. Conclusion

This paper puts light on many different methodologies and algorithms and objects of E-commerce pattern, so as a generalized view of this cumulative study shows no any system/methodology are perfect in providing solution for E-commerce so as an initiative process this paper puts forward an idea of how the input data will be preprocessed by using apriori and eclat algorithms we will generate the interesting patterns..

### 4. Acknowledgement

We would like to take this opportunity to thank our internal guide **Prof. S. P Godse**, for giving us all the help and guidance we needed. We are really grateful to them for their kind support. Their valuable suggestions were very helpful. We are also grateful to **Prof.B.B.Gite**, Head of Department of Computer Technology, **SINHGAD ACADEMY OF ENGINEERING, PUNE** for his indispensable support and suggestions.

### 5. REFERENCES


