International Journal of Circuit, Computing and Networking

E-ISSN: 2707-5931
P-ISSN: 2707-5923
IJCCN 2020; 2(1): 27-29
Received: 01-01-2021
Accepted: 03-03-2021

Jahnavi G
Department of Computer Science, SDHR College, Tirupati, Andhra Pradesh, India

Boyella Mala Konda Reddy
Assistant Professor, Department of Computer Science, SDHR College, Tirupati, Andhra Pradesh, India

Corresponding Author:
Jahnavi G
Department of Computer Science, SDHR College, Tirupati, Andhra Pradesh, India

An experimental approach for discovering association rules using FP-growth algorithm

Jahnavi G and Boyella Mala Konda Reddy

DOI: https://doi.org/10.33545/27075923.2021.v2.i1a.23

Abstract
Information mining is utilized to manage the immense size of the information put away in the data set to extricate the ideal data and information. It has different strategies for the extraction of information; affiliation rule mining is the best information mining procedure among them. It finds covered up or wanted example from huge measure of information. Among the current strategies the continuous example development (FP development) calculation is the most productive calculation in discovering the ideal affiliation rules frequent example mining is one of the dynamic examination topics in information mining. Affiliation Rule Mining is a space of information mining that spotlights on pruning up-and-comer keys. The FP-development calculation is presently probably the quickest ways to deal with continuous thing set mining. In this paper, we present a technique for mining affiliation rules utilizing FP-development calculation in enormous data sets of deals exchanges. We carry out the FP-development calculation for discovering solid affiliation rules utilizing Supermarket information, which was taken from UCI Machine Repository information. Exploratory outcomes show that this calculation can find incessant itemsets and successfully mine solid affiliation rules.

Keywords: KDD, discovering association rules using, FP-growth algorithm

1. Introduction
Knowledge discovery in databases (KDD) is defined as the non-trivial extraction of valid, implicit, potentially useful and ultimately understandable information in large databases [1]. For several years, a wide range of applications in various domains have benefited from KDD techniques and many works has been conducted on this topic. The problem of mining frequent item sets arose first as a sub-problem of mining association rules [2]. Association rule mining is one of the most important techniques of data mining. It aims to extract interesting correlations, frequent patterns, associations or casual structures among a large set of data items. A typical application is market basket analysis, which studies the buying habits of customers by searching for sets of items that are frequently purchased together [1]. Other application areas include customer segmentation, store layout, web usage mining, software defect detection, telecommunication alarm prediction, and bioinformatics.

2. Association rules
Association analysis has been broadly used in many application domains. One of the best known is the business field where the discovering of purchase patterns or associations between products is very useful for decision making and for effective marketing. A set of items is called frequent if it satisfies a minimum threshold value for support and confidence. Support shows transactions with items purchased together in a single transaction. Confidence shows transactions where the items are purchased one after the other. For frequent itemset mining method, we consider only those transactions which meet minimum threshold support and confidence requirements. Insights from these mining algorithms offer a lot of benefits, cost-cutting and improved competitive advantage.

2.1. Problem definition
Association rule mining is a data mining method to find the interesting association or correlation among a large set of data items. A formal statement of the association rule mining problem is as follows [2]. Let $\{I = I_1, I_2 \ldots I_m\}$ be a set of items. Let D be a set of transactions, where each transaction T is a set of items such that $T \subseteq I$. Associated with each transaction is a unique identifier, called TID. A transaction T contains X, a set of items in I,

if $X \subseteq T$. An association rule is an implication of the form $X \Rightarrow Y$, where $X \subset I$, $Y \subset I$ and $X \cap Y = \emptyset$. The rule $X \Rightarrow Y$ holds in the transaction set D with confidence C if C% of the transactions in D that contain X also contain Y. The rule $X \Rightarrow Y$ has support S in the transaction set D if S% of the transactions in D contain $X \cup Y$. Confidence determines the strength of the rule and support measures the frequency of the occurring pattern. A set of items is referred to as itemset. An itemset that contains k items is a k-itemset. The occurrence frequency or count of an itemset is the number of transactions that contain the itemset. If an itemset has a transaction support higher than a user-specified minimum support threshold, it is a frequent itemset. Given a set of transactions, D, the problem of association mining is to find strong rules with support and confidence greater than the given minimum support and confidence thresholds, respectively. An association rule discovery algorithm can be decomposed into two successive stages [3]. In the first stage, all sets of frequent items are discovered. In the second stage, rules are derived from these itemsets. It is important to generate all itemsets efficiently.

3. FP-Growth Algorithm

The FP-Growth Algorithm is an elective method to discover continuous itemsets without utilizing applicant ages, accordingly further developing execution. The FP-Growth Algorithm, proposed by Han in [6], is a productive and adaptable technique for mining the total arrangement of successive examples by design piece development, utilizing an all-encompassing prefix-tree structure for putting away compacted and pivotal data about incessant examples named continuous example tree (FP-tree). FP-development calculation is an effective strategy for mining all continuous itemsets without competitor age. FP-development uses a blend of the vertical and even information base design to store the data set in primary memory.

The calculation mines the continuous itemsets by utilizing a gap and-vanquish methodology as follows: FP-development first packs the data set addressing successive itemset into a regular example tree, or FP-tree, which holds the itemset affiliation data also. The subsequent stage is to separate a compacted data set into set of restrictive All hubs relate to things have a counter.

The FP-development calculation comprises of the accompanying advances

1. Scan DB once, discover incessant 1-itemset (single thing design)
2. Sort continuous things in recurrence diving request, f-list
3. Scan DB once more, develop FP-tree
4. Construct the restrictive FP tree in the succession of opposite request of F - List - produce incessant thing set

4. Experimental Results

The experiment was conducted using Weka. Weka stands for Waikato Environment for Knowledge Analysis. The software is written in the Java language and contains a GUI for interacting with data files. WEKA also provides the graphical user interface of the user and provides many facilities. WEKA is a state of-the-art facility for developing machine learning (ML) techniques and their application to real-world data mining problems. Weka implements algorithms for data pre-processing, classification, regression and clustering and association rules. It also includes visualization tools.

This section comprises the experimental analysis of Supermarket dataset was gathered from the UCI machine learning repository [13]. This dataset contains 4627 instances and 217 attributes. There are two classes of transactions i.e., Low containing 2948 records and High contains 1679 records. The summary and Statistical summary of Supermarket dataset are shown in the figure-1 and figure-2.

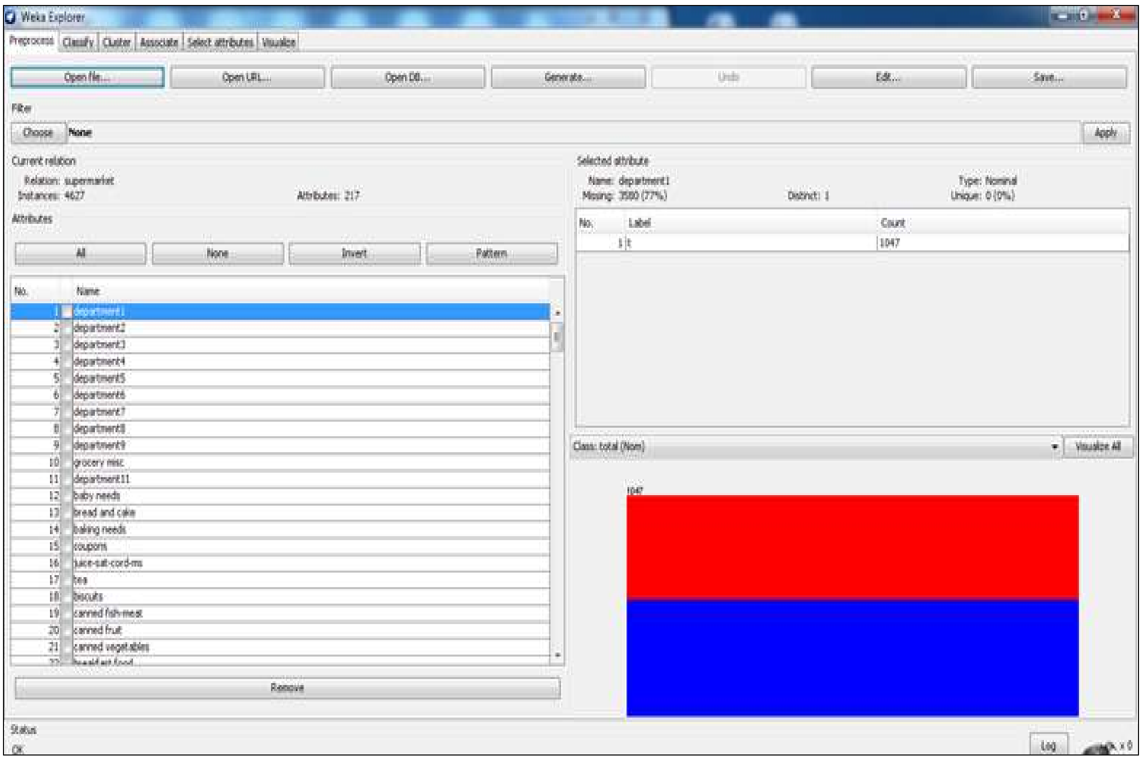

Fig 1: Summary of Dataset

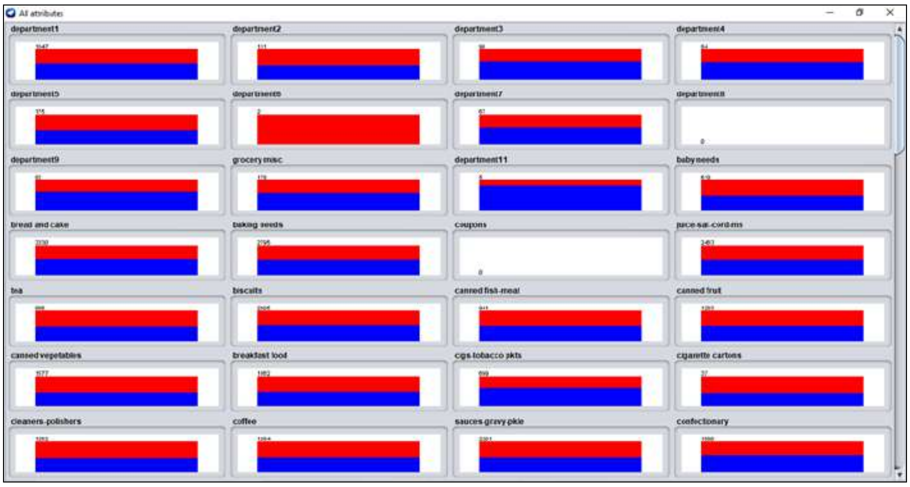

Fig 2: Statistical Summary of Dataset

4.1 Screen Shot of Apriori

Fig 3: Results of Association rules

5. Conclusion

In this paper we depicted an execution of the FP-development calculation, which contains two techniques for productively projecting a FP-tree the center activity of the FP-development calculation. This examination paper investigates and involves the regular itemset digging calculations for rules age are executed and dissected with the fitting datasets. This investigation is centered on how to track down the continuous examples proficiently utilizing FP-development calculation.

6. References

1. Agrawal R, Imieli ski T, Swami A. Mining association rules between sets of items in large databases, in Acm Sig Mod Record 1993;22:207-216.
2. Agrawal R, SRIKANT R. Fast algorithms for mining association rules. In Proc. 20th Int. Conf. Very Large Data Bases 1994, 487-499.
3. Huiying Wang, Xiangwei Liu. The Research of Improved Association Rules Mining Apriori Algorithm, FSKD, 8Th IEEE International Conference 2011;2:961-964.
4. Hu Ji-ming, Xian Xue-feng. The Research and Improvement of Apriori for association rules mining. Computer Technology and Development 2006;16(4):99-104.
5. Han J. Data Mining Concepts and Techniques, Second Edition. Morgan Kaufmann Publisher 2006, 123-134.
6. Han J, Pei H, Yin Y. Mining Frequent Patterns without Candidate Generation. In: Proc. Conf. on the Management of Data (Sigmod'00, Dallas, TX). ACM Press, New York, NY, USA 2000.
7. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/.
8. Wang Feng, Li Yong-hua, An Improved Apriori Algorithm Based on the Matrix, fbie. International Seminar on Future Bio Medical Information Engineering 2008, 152-155.
9. Ye Yanbin, Chiang Chia-Chu, A Parallel Apriori Algorithm for Frequent Itemsets Mining, sera. Fourth International Conference on Software Engineering Research, Management and Applications (SERA'06) 2006, 87-94.
10. Yu Zhen Wang. Analysis and Discussion of Web Data Mining. Development and application of computer 2003;16:72-74.