Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 7, Issue. 4, April 2018, pg.191 – 194

Study and Analysis of Infrequent Behaviour Patterns in Business Process Event Logs

S.Santhoshkumar, K.Silambarasi, E.Ramaraj
Department of Computer Science, Alagappa University, Karaikudi, Tamil Nadu, India
santhoshkumars@alagappuniversity.ac.in

Abstract— Data Mining consists of variety of techniques. Association rule mining is one of the familiar techniques used to solve many real world problems like market basket analysis, marketer prediction. The dynamic data mining is a new phenomenon used to handle event log information. It is highly used for business applications. The frequent item set mining with association rules gives better results with process log data. But least frequent items affect the performance rate of business transactions. In order reduce the noise and least significant items, filtering techniques are used. This paper is aimed to remove infrequent behavior from event logs.

Keywords— Data Mining, Association Rule Mining, Infrequent Patterns, filtering, Event logs

I. INTRODUCTION

The Data Mining technique is used in a wide range of applications. The digitization of information enabled accumulation of huge volumes of data in databases. The standard applications need data mining techniques rather dynamic business application needs a new type of technique to handle both processing and processed data. Hence log data or event logs require technique with business solutions. Process mining is a branch of process management along with data mining technique which support the analysis of business process based on log data. The term Process Mining is used in a broader setting to refer not only to techniques for discovering process models, but also techniques for business process conformance and performance analysis based on event logs. The major drawback in process mining is finding relevant information from event logs which contains overall information. The irrelevant informations are considered as noise and needed information may also incomplete or missing. For that, preprocessing techniques are used to extract the meaningful information from event logs. Frequently occurred data in event logs are considered business value and it will be used for customer profiling, behaviour analysis etc., In some cases frequently occurred itemsets will be misidentified which leads t poor results. Therefore the infrequent itemsets should be identified and eliminated before used of business applications. The primary technique used for infrequent itemset is infrequent behaviour analysis from event logs.

II. RELATED STUDY

Raffaele Conforti et al [1] proposed a process log infrequent pattern filter model using Noise-tolerant Discovery and NP complete problem solving technique. The results shown a significant improvement over fitness, precision and complexity without a negative effect on generalization. David [2] developed a new algorithm, MINIT, for finding minimal infrequent occurrent itemsets. The pruning method is also used to get better results. Kajal Nagare et al [3] proposed a weight based item set mining Eclat algorithm. The highest weight (frequently occurred) items sets are ordered sequentially. The least weighted item sets are considered infrequent and can be eliminated. Leemans et al [4] developed block-structured process models for infrequent behaviour itemsets from event logs.

III.PRELIMINARIES

A. Data Mining

Data mining [5] is the process of discovering hidden or unknown patterns in large data bases. The intersection of machine learning, statistics, and database systems. Data mining is defined as the procedure of extracting useful information from large sets of data.

B. Association Rule Mining

Association rule learning [5,6] is one of the condition based rule technique in data mining. The formation of rules based on relationships among data is used extract interesting patterns in data bases. The rules creating and interestingness of attributes in data base using support and confidence measures.

Support

Support is an indication of how frequently the itemset appears in the dataset. This can be calculated using following formula

$$Support\ (X) = \frac{|\{t \in T; X \subseteq t\}|}{|T|}$$

The support of with respect to is defined as the proportion of transactions in the dataset which contains the itemset.

Confidence

Confidence is an indication of how often the rule has been found to be true. The confidence can be calculated as

$$Confidence(X \Rightarrow Y) = Support\ (X \cup Y)/Support(X)$$

C. Frequent Itemset Mining

Frequent patterns[6,7] are item sets, subsequences, or substructures that appear in a data set with frequency no less than a user-specified threshold. Based on the support and confidence measures, the interesting patterns can be calculated and the minimum support for a variable in a transaction increases the frequent occurrence of that item. In other words, least occurrence item pairs achieve less support and those item sets are called infrequent item sets.

D. Infrequent Itemset Mining

Patterns that are often considered to be uninteresting and are eliminated using the support measure. Such patterns are known as infrequent patterns[7,8]. An infrequent pattern is an itemset or a Rule whose support is less than the minsupthreshold. Although a vast majority of infrequent patterns are uninteresting, some of them Might be useful to the analysis, particularly those that correspond to negative Correlations in data.

E. Need for Filteration

The support and confidence measures give better results basic on the interestingness between variables in database. The interesting patterns can occur frequently which achieves better accuracy, precision and least error rate also. If minimum interesting patterns gives poor performance and it produce distinct patterns. That means occurrences of item set is very less and not albe relate with existing patterns. The least items can affect the prediction rate and accuracy rate. The least occurrence item sets will affect the performance of frequent pattern

mining. Eventhough the process logs consists of repeated attributes in a transaction set, the interesting patterns are sufficiently generated.

IV. PROPOSED WORK AND EXPERIMENTAL RESULTS

The proposed work aimed to develop an application for the filtering of infrequent patterns. The event log data is used for experimentation. The first step of the work is finding support and confidence measures in giving event log files. Based on the support and confidence values of each item in a itemset and its interestingness is calculated. Based on relativity among itemset set pairs are organised. The next step of the work is removal of irrelevant or noisy data from dataset. The least instance item sets are considered as outliers or infrequent item sets. The infrequent behavior may cause the poor quality of relationships in data. The figure 1 shows the transactional data which contains item sets. The support of each item in a set is calculated.

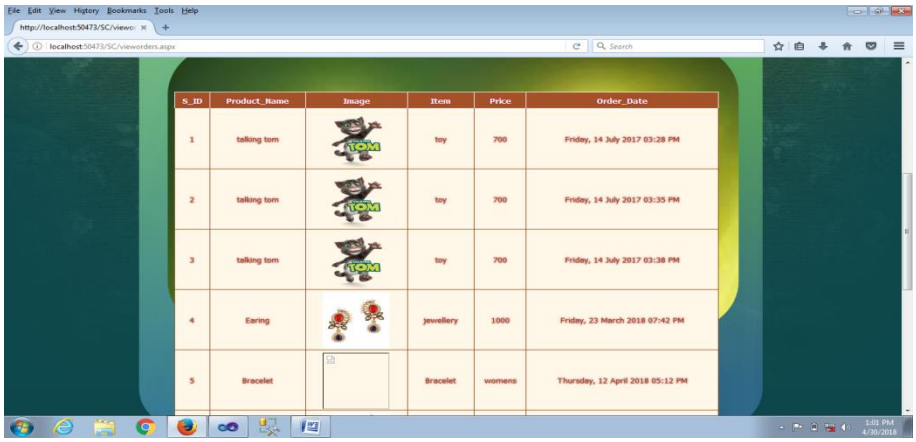

Fig 1 Transaction Data

Based on the support values the relevant items are selected along with high support valued item. The itemsets are listed with instances from highest to lowest. The least occurrence valued itemsets are having low support in a itemset. Hence the least valued itemsets are filtered using association rule mining technique. The figure 2 shows the frequent itemsets in transaction sets.

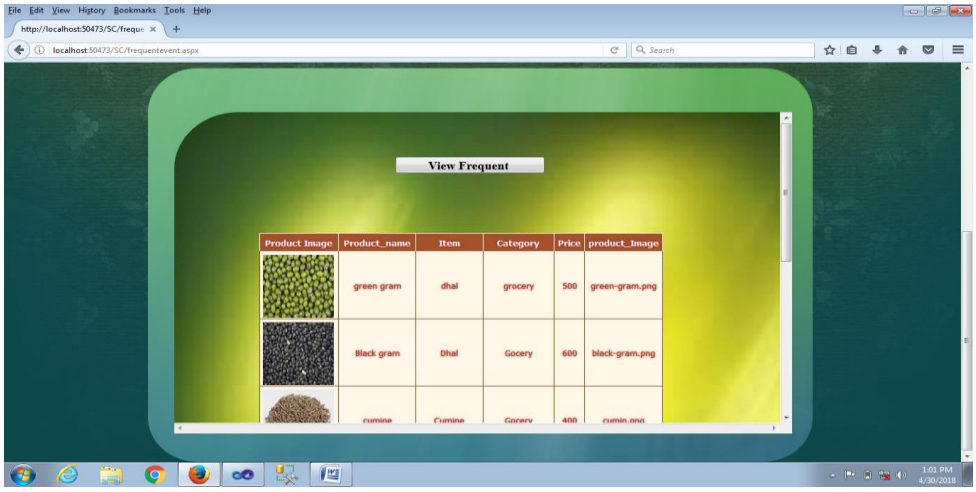

Fig 2 Frequent Itemset Prediction

The figure 3 shows the prediction of infrequent item from frequent items. The filtering of least support and confidence itemset will increase accuracy rate and reduce error rate.

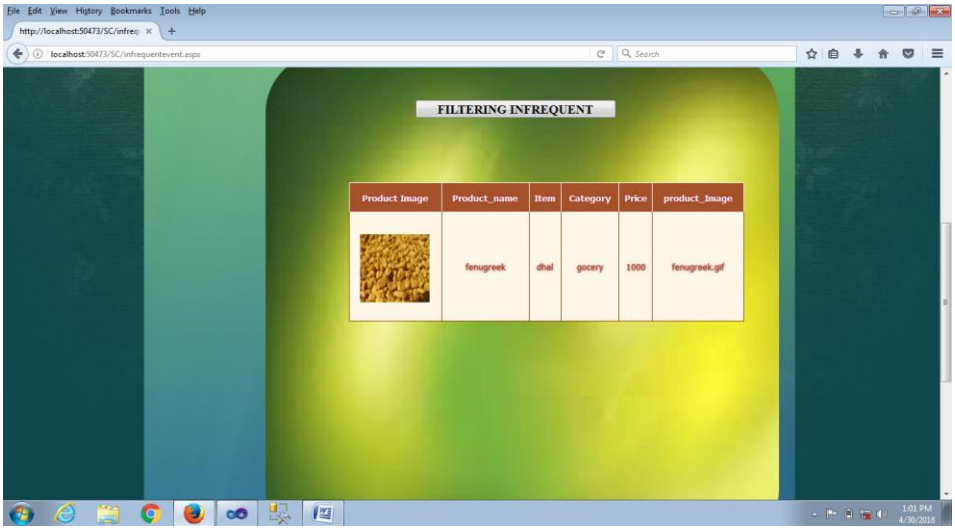

Fig 3 Filtering Infrequent Itemset

V. CONCLUSION

Data mining is a important technique which is used to solve many complex problems in real world. The process mining is a needed concept to identify frequent item set pair in a given data. The least frequent item sets are removed based on the strength of support value and confidence. This word aimed to filter infrequent items in the frequent itemsets based on support and confidence measures. The association rule mining is used to find the frequent patterns in event logs and deep learning technique is used to filter the infrequent items. The basic study and analysis of process log mining is done with sample data and results achieved better accuracy and error rate.

REFERENCES

[1] Filtering out Infrequent Behavior fromBusiness Process Event Logs Raffaele Conforti, Marcello La Rosa and Arthur H.M. ter Hofstede
[2] David J. Haglin and Anna M. Manning, On Minimal Infrequent Itemset Mining
[3] A. Adriansyah. Aligning Observed and Modeled Behaviour. PhD thesis,TechnischeUniversiteit Eindhoven, 2014
[4] W.M.P. van der Aalst. Alignment based precision checking. In Proc.of BPM Workshops, pages 137–149, 2012.
[5] Charmi Mehta, Basics of Data Mining: A Survey Paper, International Journal of Trend in Research and Development, Volume 4(2), ISSN: 2394-9333
[6] Hemlata Sahu, Shalini Shrma, Seema Gondhalakar, A Brief Overview on Data Mining Survey, International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume 1, Issue 3, ISSN 2249-6343.
[7] Adriansyah, B.F. van Dongen, and W.M.P. van der Aalst. Conformance checking using cost-based fitness analysis. In Proc. of EDOC,pages 55–64, 2011.
[8] C.C. Aggarwal. Outlier Analysis.Springer, 2013.
[9] S. Basu and M. Meckesheimer. Automatic outlier detection for time series: an application to sensor data. KAIS, 11(2):137–154, 2006.
[10] S. Budalakoti, A.N. Srivastava, and M.E. Otey. Anomaly detection and diagnosis algorithms for discrete symbol sequences with applications to airline safety. IEEE TSMCS, 39(1):101–113, Jan 2009.
[11] V. Chandola, A. Banerjee, and V. Kumar.Anomaly detection for discrete sequences: A survey. IEEE TKDE, 24(5):823–839, May 2012.

