Boolean Based Mining Algorithm for Pattern Discovery Based on Human Interaction

Ramanayagam S, Raja K, Kannan K

Abstract: ***Mining is a process that provides useful information on surfing and access pattern information based on capturing the behaviour of the user. Semantic knowledge helps to understand how the users will interact with the system. In this paper, we propose a Boolean based APriori Pattern (APP) algorithm to discover pattern based on human interaction using behavioural analysis. In the process of data mining, we have used a Boolean expression that helps to determine the pattern discovery based on the use of frequent pattern by applying association rules. The behavioural analysis is proposed based on the classification of ideas based on comments concerning positive opinion /contrary opinion during human interaction in the practical scenarios. The behavioural analysis is represented as a tree hierarchy where tree based mining is performed by the tree construction and interaction of flow patterns i.e., frequent patterns. The study shows that the successful pattern can be extracted based on the behavioural analysis of human interaction such as frequent pattern, flow interaction and relationships between the interactions.***

Index Terms**: *Semantic Knowledge, boolean, Mining, Frequent pattern, human interaction, behavioural.***

I. INTRODUCTION

Mining plays a vital role with specific to text and data in various research domains like networks, big data, machine learning, etc. Data mining play a crucial role in multiple fields like marketing and security to analyze and investigate the patterns information [1]. Nowadays, a considerable amount of data's (or) information's came into existence based on the usage of user's in all aspects. Concerning the area of research, data are extracted and classified to create certain pattern information. Here, we are specific to human interaction based on behavioural analysis to analysis the flow interaction pattern among the users. Behavioural studies reveal that the proposed idea can comment on the idea, positive/negative opinion on the comments and relationship. The above behavioural interactions are essential for the pattern of human interaction.

For the data discovery process, data mining is needed to derive the process of extraction and analyze the data based on information. Based on the data discovery, specific knowledge has to be extracted for the process of indexing using semantic knowledge [2]. Several existing algorithms are proposed to discover the pattern based on the information available with the combination of data mining and semantic knowledge. Data required for the process of information discovery based on the behavioural analysis of human interaction.

Data analysis is represented as a tree and the information has been extracted from the interaction based on the flow interaction pattern. Based on the tree-based data mining algorithm [3] [4], we have designed a mining algorithm based on tree representation to analyze the frequent pattern interaction. The flow pattern indicates the description of the frequent relationship among the interactions.

Here in this paper, our APriori Pattern (APP) algorithm follows boolean operation based on association rules to determine active pattern discovery of frequent pattern usage based on expression relationship in human interactions. The expression relations are represented in a tree hierarchy. In human intervention, i.e. behavioural analysis, we have classified the comments into positive or negative opinion during any scenarios.

In our proposed data mining process, we use semantic knowledge to index the data. Then, extracted human behaviours are patterned and discovered patterns are related together to form a flow interaction pattern, the i.e. relationship among the interaction is analyzed using boolean rules.

The rest of the paper is organized as follows, in section II, existing works are discussed. In section III, we discussed our proposed model, which is represented in tree hierarchy based on the boolean concept. In section IV, we have analyzed the performance of our proposed algorithm based on the frequency of data item taken. Finally, in section V, we have discussed the conclusion of how the proposed model behaves.

II. RELATED WORKS

Mining is one of the induction learning, which helps to extract the rules in small component and using the rules, correctness classification can be made. Problems are listed based on the decision tree on data mining algorithm as follows, property valued vacancy, multiple-valued selection and property selection criteria [5]. To resolve the problems, introduce weighted entropy into the decision tree for the proposed algorithm to improve the overall network performance [6].

To determine the data pattern based on the availability, two practical algorithms exists namely FP-growth and Apriori

Revised Manuscript Received on July 10, 2019.

Ramanayagam S, Research and Development Centre, Bharathiar University, Coimbatore, India,

Raja K, Department of Computer Science Engineering, Dhaanish Ahmed College of Engineering, Chennai, India

Kannan K, Department of Information Technology, AdhiParasakthi College of Engineering, Kalavai, India.

algorithm to classify user behaviour based on data pattern identification [7]. When compared to FP-growth Apriori algorithm is more effective in terms of network performance and time complexity. Apriori algorithm is useful in determining and identifying the frequent data pattern available for the individual datasets. The drawback of the apriori algorithm, which scans the central database every time to identify the recurring pattern as it also takes a lot of memory and execution time.

Proposed enhanced apriori algorithm limits the scans by bypassing the verification of database. Performance of the improved algorithm is calculated by reducing memory usage and execution time [8]. Propose a workflow based frequent pattern mining algorithm to solve the problem of frequent pattern discovery based on the apriori algorithm. In this persistent pattern identification, logical relations between the activities are required [9].

Application-oriented with human interaction plays a vital role in understanding the information relativity [10]. Social interaction considers behavioural analysis includes surveillance, information retrieval, medical diagnosis and computer with human interaction. Existing methods for visualizing and analyzing human interactions have many disadvantages like complex system handling, challenging to identify the negative points during the meeting and increase in multiple iterations of data. In the proposed algorithm, we have identified face to face interaction, represent the data in the form of the tree structure and data flow to determine the relationship between the interactions [11].

Proposed modified apriori algorithm to solve the problem of high time for the searching of the database regarding the frequent data item by scanning and transactions [12]. Based on the association rule, knowledge/data are discovered and analysis shows that the modified apriori algorithm has less time-consuming. A mining algorithm is proposed in which a tree represents human interaction to classify and extract the flow pattern based on communication. Social interactions are sorted based on behavioural activity to obtain frequent pattern based on semantic knowledge. In other aspects related to data mining, the data pattern is created, which helps for validation of the users [13] in networks [14].

Based on the above survey, apriori has few drawbacks and it is well suitable for mining the frequent pattern data sets. Human interaction based on behavioural analysis plays an important role in data mining.

III. PROPOSED SYSTEM

The proposed methods are associated with three elements such as interaction sets based on human interaction, interaction flow based on tree hierarchy and data mining based on boolean for pattern discovery.

A. Behavioural Interaction

In the proposed work, starting with interaction sets classified based on human interaction in a particular scenario. Interactions are represented as the human behavioural activity based on situations like a meeting, conference, etc. Various interactions are defined based on the behaviour of the human during the situation. So, we have applied behavioural based interaction sets which are represented as actions.

Set of interactions = {PROPOSE, COMMENTS, OPN, POSOPN, NEGOPN and ACK} which is explained as during interaction as POSOPN and NEGOPN are positive and negative opinion, OPN as opinion and ACK as a response are some of the suitable sets taken for my scenario as human interaction in conference/meeting.

Existing models viz. Support Vector Machine (SVM), Bayesian Network, Naïve Bayes and Decision Tree are developed to classify the behavioural activity. Here, we have obtained generated the sets of interactions applicable to this situation.

B. Interaction Flow

After classifying the interaction sets, we need to develop flow interaction between the interaction elements which are represented in a tree hierarchy. We have applied tree hierarchy which helps to create pattern discovery.

The flow contains a list of all interactions and generates relationships between the interactions.

SPON Interactions = {PROPOSE, OPN, POSOPN, NEGOPN}

REACTIVE Interactions = {ACK, COMMENTS}

The tree representations of the interactions are shown in Fig. 1 & 2 and it indicates rectangle as the interactions and arrows represents the relationship between the interactions.

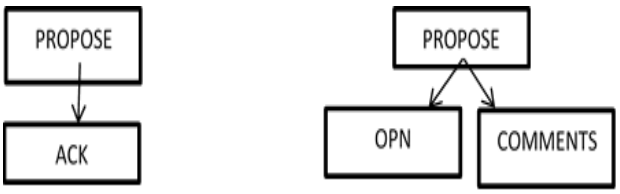

Fig. 1 Tree Representation For The Following Sets Of Interactions {PROPOSE, ACK, OPN, COMMENTS}

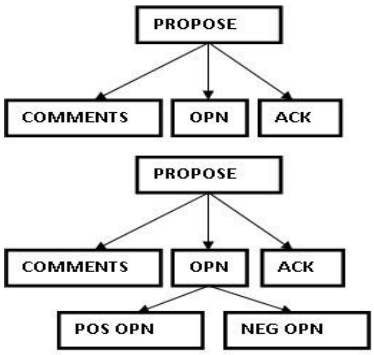

Fig. 2 Tree Representation For The Following Sets Of Interactions {PROPOSE, ACK, OPN, POSOPN, NEGOPN, COMMENTS}

In Fig. 1, PROPOSE has related two interactions such as OPN and COMMENTS but PROPOSE has ACK for the proposed idea created. No sorting is done as they spoil the direct relationship between the interactions. In Fig. 2, PROPOSE has {COMMENTS, OPN, ACK} and OPN have {POS OPN, NEG OPN}.

C. Procedure for Pattern Discovery

Based on the tree hierarchy into the flow interaction between the elements, we have applied a modified APriori Pattern Algorithm which has taken boolean based on the associated rules to calculate the pattern discovery of frequent pattern usage. Recent existing algorithms are FP-Growth and Apriori algorithm, from this apriori algorithm is giving better results based on the transactional rules with a minimal number of scans. A minimal number of scans takes less time for finding frequent pattern discovery. We propose the APP algorithm which takes less number of scans and reduced time to determining the frequent pattern discovery.

D. Algorithm

Input: Tr→ Transaction Database
S → Support
CIs → Candidate item set of size 's'
Output: To Determine the frequent item present in 's'
Fs → Frequent itemset of size 's'. // Formed based on user behaviour
Initialize the Apriori (Tr, S)
Assign CI1 →Condn(large itemset >= SthTr→Transaction)
S→ 2
While (Fs-1not equal to ϕ) // Condition to check whether it is null or not
CIs →Join(Fs-1)
For (c in CIs)
Initialize Fs not equal to zero // Transaction cannot be Zero.
Transcation id →Tid = Common Transaction IDS (c, F1)
If (|Tid| >= S then
Fs →Fs union (c)
s → s+1. // Incrementing sequence
Return Fs;
// Determine the frequent items present in's'

Following process involved in determining the frequent item present in 's' as below,

Initialize given set of transactions as Tr as represented as {PROPOSE, COMMENTS, OPN, POSOPN, NEGOPN and ACK}. We have taken four transactions with the following items from Fig. 1 & 2. Find the frequency of all the individual items based on the set of transactions. Based on the frequency, find the set of transaction ID for all the items.

Combine the set of possible items to find common transactions ids and a number of common transaction ids |Tid|. Based on the above process, we can determine frequent items present in size's'. Frequent item calculation for APriori Pattern (APP) Algorithm helps to reduce the number of scans taken for this process.

Tree representation is used for finding only the individual transaction sets and with the transaction sets and APP algorithm, we can determine the frequent items present in 's'. Based on the frequent itemsets and transaction count, we can find which tree is applied frequently from Fig. 1 & 2.

IV. PERFORMANCE ANALYSIS

In the performance analysis, taken frequent itemsets based on scans and time period which helps to determine effective pattern discovery of frequent usage of pattern based on human interactions.

In Fig. 3, compared various frequent item sets with the number of scans taken to complete the process. We inferred that the APP algorithm takes a minimal number of scans when compared to existing algorithms like traditional apriori algorithm [15] and enhanced apriori algorithm [16]. Assume initial weights and biases are:

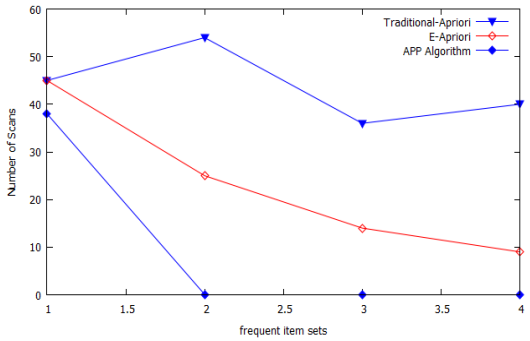

Fig. 3 Frequent Itemsets Vs Number Of Scans

In Fig. 4, we have compared various frequent item sets with the time taken to complete the process. We inferred that the APP algorithm takes a minimal number of amount of time when compared to existing algorithms like traditional apriori algorithm and enhanced apriori algorithm.

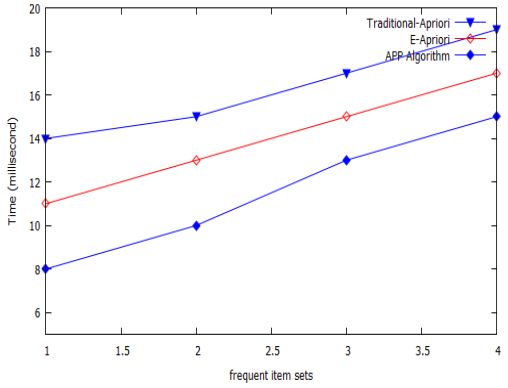

Fig.4 Frequent item sets Vs Time (Millisecond)

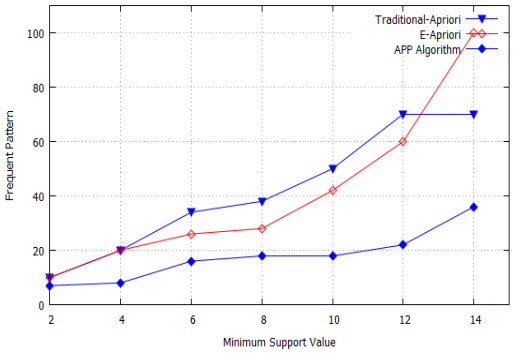

Fig 5. Minimum Support Value Vs Frequent Pattern

In Fig. 5, we have considered two parameters such as

minimum support value and frequent pattern to analyses whether the proposed APP algorithm is effective when

compared to the other existing algorithms like traditional Apriori and E-Apriori. Here we have inferred that the proposed algorithm used the least frequent pattern based on the variation in the minimum support value. The accuracy percentage is measured based on the variation in the number of sessions which is represented in Fig. 6. Fig. 6 conclude that the proposed APP algorithm performs better when compared to the existing ones in terms of accuracy and accuracy is achieved better in all variation in the session parameter.

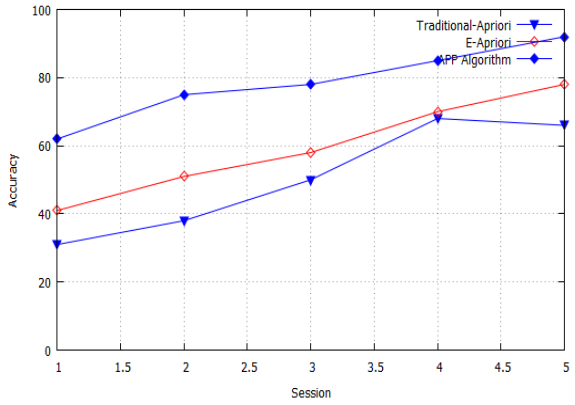

Fig. 6 Session Vs Accuracy

V. CONCLUSION

In this article, boolean based mining algorithm with the flavour of apriori to determine the frequent data pattern based on human interaction has been proposed. Based on the behavioural analysis, the sets of interactions {{PROPOSE, COMMENTS, OPN, POSOPN, NEGOPN and ACK} are taken. It has been noted that the interaction flows are represented as a tree hierarchy. The tree-based mining algorithm is proposed in the paper to perform and analyze the construction of the tree and its related interactions to form the flow pattern i.e. frequent pattern. In this paper, analyzed the frequent data items based on a number of scans and time taken to execute the process for the proposed algorithm. Based on the parameters discussed, and inferred that the proposed algorithm is efficient than the existing ones in terms of efficiency, accuracy and throughput.

REFERENCES

1. Andreas Holzinger, "On interaction in Data Mining," In Proc. of International Conference on Computer Vision Theory and Applications (VISAPP), 2014.
2. K. Rajaraman and A.-H. Tan, "Mining semantic networks for knowledge discovery," In Proc. 3rd IEEE International Conference on of Data Mining, ICDM, 2003, Doi: 10.1109/ICDM.2003.1250995.
3. Zhiwen Yu, Zhiyong Yu, Xingshe Zhou, Christian Becker and Yuichi Nakamura, "Tree-Based Mining for Discovering Patterns of Human Interaction in Meetings," IEEE Transactions on Knowledge And Data Engineering, Vol. 24, No. 4, April 2012.
4. AsmitaShejale and Vishal Gnagawane, "An implementation of efficient techniques for tree-based mining in human social dynamics," In Proc. International Conference on Data Mining and Advanced Computing (SAPIENCE), 2016, Doi: 10.1 109/SAPIENCE.2016.7684138.
5. Libina Rose Sebastian, SheebaBabu and Jubilant J Kizhakkethottam, "Challenges with big data mining: A review," In Proc. of International Conference on Soft-Computing and Networks Security (ICSNS), 2015.
6. Linna Li and Xuemin Zhang, "Study of data mining algorithm based on decision tree," In Proc. International Conference on Computer Design and Applications (ICCDA), 2010, Doi: 10.1109/ ICCDA.2010.5541172.
7. K. Dharmaraajan and M. A. Dorairangaswamy, "Analysis of FP-growth and Apriori algorithms on pattern discovery from weblog data," In Proc. IEEE International Conference on of Advances in Computer Applications (ICACA), 2016, Doi: 10.1109/ICACA .2016.7887945.
8. S. P. Aditya, M. Hemanth, C. K. Lakshmikanth and K. R. Suneetha, "Effective algorithm for frequent pattern mining," In Proc. of International Conference on IoT and Application (ICIOT), 2017, Doi: 10.1109/ICIOTA.2017.8073615.
9. Weidong Zhao and Mao Ye, "A Workflow Frequent Pattern Mining Algorithm," In Proc. of International Conference on Wireless Communications, Networking and Mobile Computing, WiCom 2007, Doi: 10.1109/WICOM.2007.911.
10. R.S. Parpinelli, H.S. Lopes and A.A. Freitas, "Data mining with an ant colony optimization algorithm," IEEE Transaction on Evolutionary Computation, Vol. 6, Issue. 4, pp. 321-332, 2002.
11. G. Vinitha Sanchez and T.S. Vishnu Priya, "System Based Mining For Discovering Human Interaction In Meetings," International Research Journal of Engineering and Technology (IRJET), Vol. 4, Issue. 8, pp. 1891-1893, 2017.
12. Mohammed Al-Maolegiand BassamArkok, "An Improved Apriori Algorithm For Association Rules," International Journal on Natural Language Computing (IJNLC) Vol. 3, No.1, pp. 21- 29, 2014.
13. Rajakumar Arul, Gunasekaran Raja, KottilingamKottursamy, Sathya Pavithra, Swaminathan Venkatraman, "User Path Prediction Based Key Caching and Authentication Mechanism for Broadband Wireless Networks," in the Springer – Journal on Wireless Personal Communications, Vol.94, Issue 4, pp.2645 – 2664, June 2017.
14. Ramkumar Jayaraman, Gunasekaran Raja, Dipak Ghosal, Rajakumar Arul and Sabareesh Kumar A, "A Compatibility Vector Technique for Cooperative Scheduling and Channel Assignment Algorithm in Broadband Wireless Networks," in ACM / Springer - Journal of Mobile Networks and Applications (MONETS), Vol. 22, Issue. 4, pp. 730-742, Mar. 2017.
15. Zhiwen Yu, Zhiyong Yu, Xingshe Zhou, Christian Becker and Yuichi Nakamura, "Tree-Based Mining for Discovering Patterns of Human Interaction in Meetings," IEEE Transactions on Knowledge and Data Engineering, Vol. 24, No. 4, April 2012.
16. Sudhir Tirumalasetty, ArunaJadda and Sreenivasa Reddy Edara, "An Enhanced Apriori Algorithm for Discovering Frequent Patterns with Optimal Number of Scans," International Journal of Computer Science, 2015, arXiv:1506.07087.

AUTHORS PROFILE

Ramanayagam S is an Associate Professor in the MCA department, Hindustan University, Chennai. He is a Research Scholar in Bharathiar University. He has 28 years Academic experience. He has received MCA from Bharathidasan University, Trichy. He has completed his MSc in Annamalai University, Chidambaram. He has presented papers in 3 International Conferences. His Area of interests are Data Mining, Big Data.

Dr. K. Raja is working as Principal at Dhaansih Ahmed College of Engineering, Chennai, India, since Sep 2016. He completed his B. Sc (Mathematics) from Madras University,1989, B.E (Computer Science and Engineering) M. E (Computer Science and Engineering) from Madras University, 2001, Ph. D degree in Knowledge-based systems from Sathyabama University, India, 2006. He is a life member of various professional bodies like Institution of Engineers India, Computer Society of India, and Indian Society for Technical Education, International Association of Computer Science and Information Technology etc. He has more than two decades of teaching experience. He has published research papers in peer-reviewed, reputed 45 International Journals, 8 National Journals, 54 International (Singapore, Malaysia, Sri Lanka, Doha Qatar, and India), and National conferences. He has written 5 books in various fields.

Dr. K. Kannan is presently working as an Associate Professor in Department of IT, Adhiparasakthi College of Engineering, Kalavai, Tamilnadu, India, since July 2008. He has received his Ph.D. from Sathyabama University. He has completed his M. E., (CSE) from Anna University, Chennai, India, 2008. He is a Life Member in professional bodies like ISTE –

Indian Society for Technical Education, and Professional Member of ACM - Association for Computing Machinery.
He has earned 19 Years of Academic Experiences from various reputed educational Institutions. He has published research papers in peer-reviewed, reputed 7 International Journals, 6 National Journal, and presented the papers in 25 International (Singapore, Doha Qatar, and India), and National Conferences. His areas of interest are Knowledge and Data Engineering, Decision Support Systems, Service Oriented Architecture, Web Technology Data Structures, and Data Warehousing.