A Survey Paper on Discovering Patterns from Human Interactions

Samreen Sadaf Qazi
ME(Computer)-II
Vishwakarma Institute Of IT, Pune

Prof. Leena A. Deshpande
Assistant Professor
Vishwakarma Institute Of IT, Pune

Abstract- Discovering knowledge from human interactions of any group activity is important for interpreting and understanding the behavior pattern of human being in meetings, interviews, group discussion and so on.Finding frequent patterns from human interactions can play an important role in the fields like finalizing business deals, interpreting the human interactions and behavior in meetings, facing interviews, getting patients' detail to diagnose patients, and in setting winning strategies for games. Interaction indicates the intention of a person towards the topic in discussion. Discovered patterns can be used in meetings to determine the frequent interactions, and relationship between interactions. They can also be used in sports to identify correlations among play patterns, to improve the performance of playing teams, and to make winning strategies for future.

Keywords: Human interaction, frequent pattern mining, interaction pattern, association rule mining

I. INTRODUCTION

Data mining is a powerful method of extracting knowledge/information from large amount of data. It has great attention in recent years because of growing amount of data and the persistent need of turning the huge data into information. It is widely adapted in many fields like bioinformatics, business analytics, marketing, security and many more.

An important task of data mining is discovering frequent patterns that play an important role in clustering, associations mining, correlations, etc. Frequent patterns are patterns that appear in a data set frequently, such as itemsets, subsequences, or substructures. An itemset A (or subsequence, or substructure) is said frequent if it satisfies the predetermined minimum support count, where

$$\text{Supp}(A) = \frac{no.\,of\ occurrences\ of\ A}{total\ no.\,of\ itemsets\ in\ a\ dataset}$$

An association rule is an implication of the form A→B, A,B are itemsets, and A∩B=Ø. The rule A→B from dataset D holds the minimum support count of AUB in D. this rule has confidence c in the dataset D.

$$confidence(A \rightarrow B) = P(B/A)$$

Association rule mining, in general can be viewed as a two step process.

- Find all frequent itemsets
- Generate strong association rules from the frequent itemsets.

A. Human Interactions:

A human interaction is one of the most important characteristics of any group activity. Group activity could be anything like meeting, group discussion, debate, presentation, sports, etc. Research in such field may include: gesture, gaze, and probably other modalities, over and above it, the speech we usually associate with human-human communication. From the technological perspective, this human style of communication poses lots of challenges for researchers.

Meetings are an important communication and coordination activity of teams: status is discussed, new decisions are made, alternatives are considered, details of topic are explained, information is presented, and new ideas are generated. As such, meetings contain a large amount of rich project information that is often not formally and properly documented. Capturing all of this meeting information has been a topic of research in several communities over the past decade.

B. Objectives:

Objective of this paper is to study various pattern mining algorithms that can be used for discovering knowledge (patterns) from human interactions. And to study and compare various association rule mining algorithms.

The rest of the paper is organized as: Section 2 describes the various pattern mining techniques evolved. Section 3 describes the pattern mining from human interactions. Section 4 contains comparative study of various association rule mining algorithms. Finally, the paper is concluded in section 5.

II. FREQUENT PATTERN MINING

Frequent pattern can be an itemset, subsequence, or substructure. Mining frequent tree pattern is an important open research area. Previous research studies highly suggest the pattern growth method for efficient pattern mining. Authors of [1] have developed a pattern growth method for mining frequent tree patterns. Two algorithms, Chopper and XSpanner, have been devised, from which XSpanner algorithm is faster than Chopper. These two algorithms perform better than TreeMinerV of M.J.Zaki, Efficiently mining frequent trees in a forest of KDD'02. In that, Chopper consists of two separate phases (i) mining sequential patterns and (ii) the extraction of frequent tree patterns. It generates and tests all possible tree patterns of the database. XSpanner algorithm combines these two phases of Chopper algorithm.

In paper [2], the problem of mining sub-trees in a forest(database) of rooted, labeled and ordered tree is formulated. And presented TREEMINER algorithm to discover all frequent subtrees in a forest, using a new data structure called scope-list. Tree mining algorithm is also applied to analyse RNA structure and phylogenetic data sets from bioinformatics domain. Authors of [2] have developed a framework for non-redundant pattern generation. There work allows one to discover all subtrees in a forest, as well as all subtrees in a single large tree.

A mining method to extract the important patterns of group activity for multimodal interactions is proposed in paper [3]. It extracts simultaneous occurring patterns of events in interaction such as gaze and speech. It uses wearable and ubiquitous sensor client systems to record human interactions. Authors have not developed a way to determine that how many patterns should be extracted, it has to be decided by the designer. Also it deals with simultaneous occurring patterns and not with sequential patterns.

Effective use and updating of the discovered patterns is an open research area. Therefore, authors of [7] have presented an effective pattern discovery technique that also includes the process of pattern evolving and pattern deploying. This technique of [7] of pattern discovery considers the negative training data to remove the noisy and ambiguous patterns which is called as evolution of patterns.

Tree based pattern mining algorithm is generally used in analyzing tree structures and in extracting the interaction patterns. Tree-based pattern mining algorithm is suggested in paper [8] and is used for finding patterns of human interactions in meeting. This work is extended in [12] and the authors of [12] have expanded Graph-based substructure pattern mining with user interaction flow with clustering.

The above frequent pattern mining algorithms may often generate an enormous number of frequent patterns that may contain a lot of redundancy. And this can create a challenge on understanding, visualizing and analyzing the generated patterns. Therefore, authors of [15] have proposed an algorithm called MinRPset to generate minimum

representative patterns that best approximate the all other patterns. It produces the smallest solution for a given problem. But MinRPset is very time-consuming and space-consuming too. So, authors have developed another algorithm to overcome this problem and proposed FlexRPset which allows users to trade-off between efficiency and result-size.

III. MINING PATTERNS FROM HUMAN INTERACTIONS

In paper [5] Authors have identified the 10 challenging problems in data mining and machine learning for future research in these fields. Sequential and time series data mining is one of the important problems. Clustering, classification, and trend prediction of these data is an important open topic.

These 10 challenging problems are:

- Developing a Unifying Theory of Data Mining
- Scaling Up for High Dimensional Data and High Speed Data Streams
- Mining Sequence Data and Time Series Data
- Mining Complex Knowledge from Complex Data
- Data Mining in a Network Setting
- Distributed Data Mining and Mining Multi-Agent Data
- Data Mining for Biological and Environmental Problems
- Data Mining Process-Related Problems
- Security, Privacy, and Data Integrity
- Dealing with Non-Static, Unbalanced and Cost-Sensitive Data

Recognition of any human interactions in group activities is an important step towards understanding of interactions without or with minimal manual work. It has proved to be useful for research purpose in the fields of image or speech processing and human-computer interaction. Numerous work has been done in finding human behavior patterns using random techniques.

Now a days, meeting data, sports data is increasing immensely. Because of that analysis of such data is getting more importance. Finding frequent patterns from human interactions can play an important role in the fields like finalizing business deals, interpreting the human interactions and behavior in meetings, facing interviews, getting patients' detail to diagnose patients, and in setting winning strategies for games. Cognitive researchers could use the results of analysis in their domain.

Discovering meaningful information from interaction of any group activity record is difficult because these records do not contain any structural information.

In paper[4] authors have surveyed and discussed various ways of indexing meeting records by categorizing existing approaches along multiple dimensions. And then introduced the notion of creating indices based upon user interaction with domain-specific artefacts. Where created index was a time-point which describes a point in a meeting record.

Paper [9] is a step towards learning structure of human interactions in TV shows. Objective of this paper is recognition and spatiotemporal localization of two-person interactions in video for automatic understanding of a scene. It deals with four interaction classes: handshakes, high fives, hugs, and kisses. It addressed the challenges: (i) irrelevant variation, and (ii) avoiding learning background noise when codifying the interaction. First it locates people to focus on regions where interaction could happen i.e. reducing the search space of an interaction. Localization of people is done by upper body detector which is linked by a combination of KLT tracking and clique partitioning clustering to form tracks. Multiclass linear Support Vector Machine is used for head orientation classification. It also handles asymmetrical interactions.

In paper [14], authors have used mining algorithm for product session data for improving efficiencies. They have proposed fuzzy mining algorithm which transforms every quantitative value into fuzzy set which includes linguistic terms. Then it calculates scalar cardinality of every linguistic term for all transaction data. After that fuzzy association rules are mined.

Paper [8] focuses discovering higher-level knowledge about human interaction from perspective of data mining. Interaction tree pattern mining algorithm is designed to analyze tree structures and extract interaction flow patterns. Authors of [8] have adopted a multimodal method to infer human interactions based on various features like gestures, attention, speech tone, and speaking time, interaction occasion, and information about the previous interactions. To infer the type each interaction four classification models, Support Vector Machine, Bayesian Net, Naïve Bayes, and Decision Tree, ware used. In that, SVM outperforms the others having recognition rate of approximately 80 percent.

IV. ASSOCIATION RULE MINING FOR GROUP ACTIVITIES

Association rule mining algorithms are used to extract the associations among large amount of items or events. It is extensively used to discover the patterns of customers' buying behavior in retail. It can be used to find human interaction pattern.

Association rule mining can be used in analyzing the sports data. Because real time cricket data is too complex and rapidly growing in size which far exceeds the human abilities to analyze.

Paper [6] is aimed to find the association rules for sports data using Principal Component Analysis. Major intention of authors is to model an automated framework to identify specifics and correlations among play patterns which can be further represented as useful information to improve or modify coaching strategies and methodologies to confine the team/ individual performance. When the size and dimension of annotated database is large, PCA proceeds as a compression mechanism. Since PCA serves as a mathematical model to reveal the relationship between two more variables, the interrelationships between patterns can easily be identified. PCA focuses on the variation parameters, only those patterns which are close to the centre are taken as frequent patterns and others are left as infrequent. Therefore, PCA optimizes the association rule mining tasks.

In another paper [10] association rule mining is used to mine relationship from sports data on performance of Indian Cricket team. Association among factors such as outcome of toss, batting first or batting second, playing in a home ground or playing abroad, and the outcome of match, i.e. win or loss is examined in this paper. Result of this association rule mining is that the performance of Indian cricket team in last ten years is encouraging as compared to entire period. The work of this paper is limited to the matches played by Indian cricket team. And the article has not considered the direct impact of team composition on the match outcome.

Table : Comparison of various Association Rule Mining Algorithms [13,17]

Algorithm	Advantages	Disadvantages	Application
AIS	Focuses on improving quality of database and process the decision support queries.	Candidate sets generated on the fly.	Not frequently used, but when used is used for small problems.
	Easy to use	Size of candidate set large.	
	Better than SETM	Requires multiple scans on whole database	
	-	Needs more memory	
SETM	Separates generation from counting.	Very large execution time.	Not frequently used
		Size of candidate set large.	
Apriori	Fast, More efficient than AIS	Takes a lot of memory.	Best for closed item sets
	Less candidate sets. Generates candidate sets from only those items that were found large.		
Apriori TID	Doesn't use whole database to count candidate sets	-	Used for smaller problems

	Better than SETM, Better than Apriori for small databases		
	Fast		
Apriori Hybrid	Better than both Apriori and AprioriTID	Used where Apriori and AprioriTID used	-
FP-Growth	Only 2 passes of dataset	Using tree structure creates complexity Not suitable for incremental mining	Used in cases of large problems as it doesn't require generation of candidate sets
	Compresses data set		
	No candidate set generation required so better than Apriori		
Rapid Association Rule Mining	Avoids candidate generation process	Requires more memory	-
	Faster than FP-Tree algorithm		

V. CONCLUSION

This paper presents study of discovering patterns from human interactions in group activity. This includes study of various techniques used for finding frequent patterns with various methods evolved in pattern mining technique along with the comparison of different association rule mining algorithms. In future, patterns can be mined from human interactions of different group activities like panel discussion, debated, sports meeting etc.

REFERENCES

[1] C. Wang, M. Hong, J. Pei, H. Zhou, W. Wang, and B. Shi, "Efficient Pattern-Growth Methods for Frequent Tree Pattern Mining", *Proc. Pacific-Asia Conf. Knowledge Discovery and Data Mining (PAKDD '04)*, pp. 441-451, 2004.

[2] M.J. Zaki, "Efficiently Mining Frequent Trees in a Forest: Algorithms and Applications", *IEEE Trans. Knowledge and Data Eng.*, vol. 17, no. 8, pp. 1021-1035, Aug. 2005.

[3] Tomoyuki Morita, Yasushi Hirano , YasuyukiSumi, "A Pattern Mining Method for Interpretation of Interaction", *ICMI'05, ACM 2005*, October 4–6, 2005, Trento, Italy, 1-59593-028-0/05/0010.

[4] W. Geyer, H. Richter, and G.D. Abowd, "Towards a Smarter Meeting Record—Capture and Access of Meetings Revisited", *Multimedia Tools and Applications*, vol. 27, no. 3, pp. 393-410, 2005.

[5] Q. Yang and X. Wu, "10 Challenging Problems in Data Mining Research", *Int'l J. Information Technology and Decision Making*, vol. 5, no. 4, pp. 597-604, 2006.

[6] P. UmaMaheswari, Dr. M. Rajaram, "A Novel Approach for Mining Association Rules on Sports Data using Principal Component Analysis: For Cricket match perspective", *IEEE International Advance Computing Conference (IACC 2009) Patiala, India*, 6-7 March 2009.

[7] N. Zhong, Y. Li, Sheng-Tang W. "Effective Pattern Discovery for Text Mining", *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 1, January 2012.

[8] Zhiwen Yu, Zhiyong Yu, Xingshe Zhou, Christian Becker, 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.

[9] Alonso Patron-Perez, Ian Reid, "Structured Learning of Human Interactions in TV Shows", *IEEE Transactions On Pattern Analysis And Machine Intelligence*, VOL. 34, NO. 12, DECEMBER 2012.

[10] K. Antony ArokiaDurai Raj, Panchapakesan Padma , "Application of Association Rule Mining: A Case Study on Team India",*2013 International Conference on Computer Communication and Informatics (ICCCI -2013)*, Jan. 04 – 06, 2013, Coimbatore, INDIA.

[11] Tulips Angel Thankachan, Dr.KumudhaRaimond, "A Survey on Classification and Rule Extraction Techniques for Datamining", *IOSR Journal of Computer Engineering (IOSR-JCE)*, e-ISSN: 2278-0661, p- ISSN: 2278-8727, Volume 8, Issue 5, Jan. - Feb. 2013, PP 75-78.

[12] D. Karthika,R. RangaRaj, "A Graph-Based Interaction Pattern Discovery for Human Meetings", *International Journal of Advances in Computer Science and Technology*,Volume 2, No.8, August 2013, ISSN 2320 – 2602.

[13] T. Karthikeyan, N. Ravikumar, "A Survey on Association Rule Mining", *International Journal of Advanced Research in Computer and Communication Engineering*,Vol. 3, Issue 1, January 2014.

[14] Nikhil Gaikwad, Gaurav Chavan, Twinkle Samal, Avinash Sonule, "Implementation of Mining Algorithms for Improving Efficiencies in Product Session Data", *International Conference on Engineering Technology and Science-(ICETS'14)*, Volume 3, Special Issue 1, February 2014.

[15] Guimei Liu, Haojun Zhang, and Limsoon Wong, "A Flexible Approach to Finding Representative Pattern Sets", *IEEE Transactions On Knowledge And Data Engineering*, Vol. 26, no. 7, July 2014.

[16] Jiawei Han, Micheline Kamber, "Data Mining- Concepts and Techniques", *Elsevier Inc.*, Indian ISBN: 978-81-312-0535-8, 2006.

[17] Manisha Girotra, Kanika Nagpal, Saloni Minocha, Neha Sharma, "Comparative Survey on Association Rule Mining Algorithms", *International Journal of Computer Applications (0975 – 8887)* Volume 84 – No 10, December 2013.