RESEARCH ARTICLE OPEN ACCESS

Tree Based Mining for Discovering Patterns of Human Interactions in Meetings

Asmita Shejale, Vishal Gangawane
Department of CE & IT, VJTI, Mumbai, India
Department of CE &IT, VJTI, Mumbai, India

Abstract
Meetings are an integral part of workplace dynamics also an important communication and coordination activity of teams: statuses are discussed, decisions, alternatives are considered, details are explained and ideas are generated. In this work, data mining techniques to detect and analyze frequent interaction patterns. We look forward to discover various types of new knowledge on interactions. An interaction tree pattern mining algorithms was proposed to analyze tree structures and extract interaction flow patterns. In this paper we propose the tree based mining for human interaction flow in a discussion session is represented as a tree. In this work we extend an interaction tree mining algorithm in three ways. First, we propose a mining method to extract frequent patterns of human interaction. Second, we explore embedded sub tree mining for hidden interaction pattern discovery. Third, we propose temporal data mining techniques for extracting the temporal patterns from the captured content of time series of different meetings in particular time periods such as month or year. Because of the human integration activities varied based on time and experience of events. For extracting temporal pattern mining we use hidden markov model (HMM) along with tree mining algorithm.

Index Terms— Human Interaction, Tree-Based Mining, Embedded Subtree Mining, Hidden Markov Model, Temporal Tree Mining

I. INTRODUCTION

In the social dynamics, human interaction is the one of the important for understanding how a human's reactions or human activities under the meetings. And determining whether the meeting was well organized or not efficient. Because it is the one of the main issue in the meetings. Several methods have been used to found the interaction of the flow in the meeting at each human[2]. The human interaction flow or the human interaction is the sequence of communications, such as proposing an idea, giving comments, expressing an opinion, etc., between the participants of the meeting. It is used to know the user role, attitude, or intention toward a topic and their suggestion about the meeting, requesting information[1]. To understand the human interactions and interference of the human interactions in meetings, first we need to discover higher level semantic knowledge called interactions flow often occur in a discussion. It encompasses what interaction flow discussion usually follows, and relationships between the exist among interactions. This knowledge will help to describes important patterns of interaction[2]. Meetings constitutes the natural and essential cases in the people interaction, becomes challenging problem for several conditions and a relatively well-defined dictionary of relevant actions. The previous work of the paper investigates to discover patterns can be utilized to evaluate whether a meeting discussion is efficient and to compare two meeting discussions using interaction flow as a key feature.

Discovering semantic knowledge is significant for understanding and interpreting how people communicate in a meeting discussion. As such, meetings contain a large amount of rich project information that is often not formally documented. Capturing all of this informal meeting information has been research over communities over the past decade. The most common way to capture meeting information is through notes-taking. However, fully writing down the content of a meeting is a difficult one. And it can result in an inability to both take notes at the same time participate in the meeting. Many data mining problems can be represented with non- linear data structures like trees. In the previous work of this paper, consider data mining techniques to detect and analyze frequent interaction patterns. So the propose work of this paper, hope to discover various types of new knowledge on interactions in the meetings at every participant in a meeting[2]. Tree-based mining, Interaction tree pattern mining algorithms examine tree structures and extract interaction flow patterns. Mining human interactions is important for accessing and understanding meeting content. The Previous work of the tree based mining method discovers the human interaction in the only one dataset [3]. Various categories of the datasets were not considered. We will develop several applications based on the discovered patterns in human meetings. So we propose our work with the various categories of the meetings. We extend our previous work in three ways. In first method, we propose the mining methods that extract frequent patterns of human interaction for various categories of meetings. Second we propose our work with the embedded subtree mining. Finally we propose the temporal tree mining algorithm with the time period series for temporal patterns.

A. Data Mining

Data mining is a powerful method for discovering new knowledge, which is adopted in many fields, such as supermarket (retail), Banking services and Medical patient histories, etc., Alternative term for data mining is knowledge mining, knowledge extraction and data pattern analysis. Knowledge discovery in databases process, or KDD is relatively is used to guide search or the process of discovering new patterns from large datasets. Pattern represents knowledge if it is easily understand by humans. Measures of pattern can be used to guide the discovery process. The goal of data mining is to extract knowledge from a data set in human-understandable information. Data mining is the entire process of applying technologies, including new techniques for knowledge discovery, from dataset. Databases, Text Documents, Computer Simulations, and Social Networks are the Sources of Data for Mining.

B. Data Mining in Human Interactions

In this paper, we study data mining techniques to detect and analyze frequent interaction patterns. And hope to discover various types of new knowledge on interactions. Human communication flow in a discussion session is represented as a tree. Tree-based mining, designed interaction tree pattern mining algorithms for constructed tree datasets. It is used to analyze tree structures and extract interaction flow patterns from the tree dataset. An interaction flow that appears frequently reveals relationships between different types of interactions. The tree-based interaction pattern mining method is used to mine the frequent interactions. A tree is used to represent an interaction flow in a session[1]. It is an acyclic connected graph, also rooted, directed, and labeled. There would be some differences in the frequent interaction patterns for different meeting styles. We survey embedded tree mining for hidden interaction pattern discovery. Third, we propose temporal data mining techniques for extracting the temporal patterns from the captured content of time series of different meetings in particular time periods[6][8].

C. Tree Based Mining

Finding frequent item sets in databases are the fundamental operation of association rule mining. Mining frequent tree patterns have many useful applications in XML mining, marketing, banking, network routing. We propose a mining method to extract frequent items of human interaction based mining on the captured content of meetings. Human interactions, such as proposing an idea, giving comments, opinions are constructed as a tree.

Tree-based interaction mining algorithms are designed to analyze the structures of the trees and to extract frequent interaction flow patterns in a tree dataset[3]. Embedded sub trees are a subtree, which is similar to subtree mining but it allows not only mining at direct parent child traversal. It is for ancestor descendent subtree mining. Capturing all of this informal meeting information is omitted by using tree based mining approach. A mining method to extract frequent patterns of human interaction based on the captured content is of human participated meetings. The mining results can be used for indexing meeting semantics also existing meeting capture systems could use this technique as a smarter indexing tool [4][5]to search and access particular semantics of the meetings. Interaction tree pattern mining algorithms to analyze tree structures and extract interaction flow patterns in different ways. Embedded tree mining for hidden interaction pattern discovery for hidden pattern discovery. For extracting temporal pattern mining we use hidden Markov model along with tree mining algorithm [2].

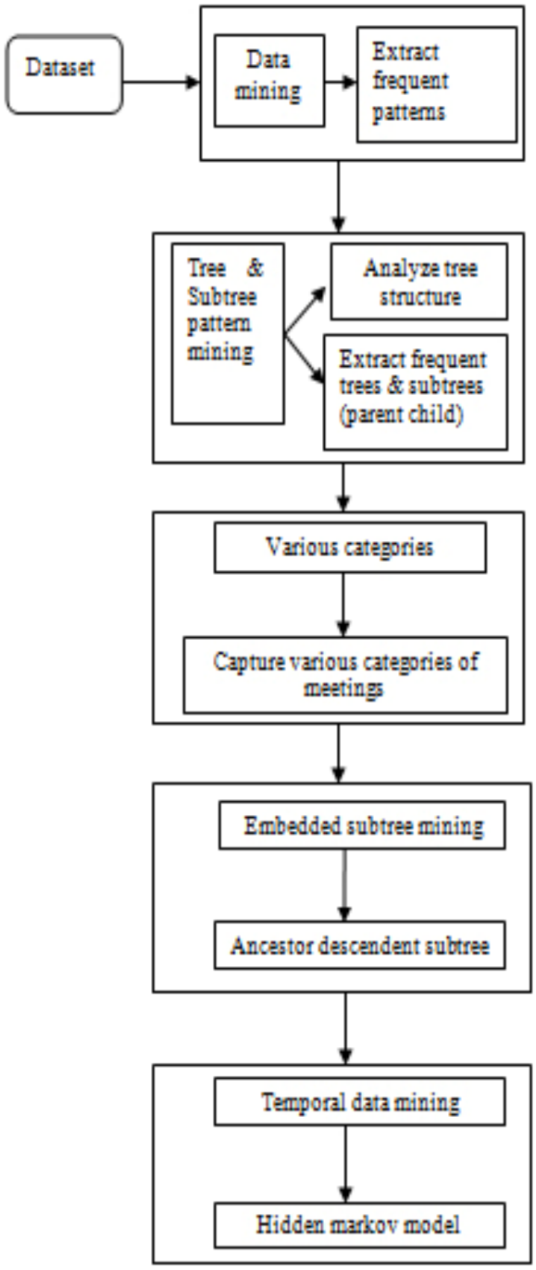

Fig. 1: Architecture Diagram for Discovering Patterns of Human Interactions in Meetings.

II. PROBLEM DEFINITION

Discovering semantic knowledge is significant for understanding and interpreting how people interact in a meeting discussion. As such, meetings contain a large amount of rich project information that is often not formally documented. Capturing all of this informal meeting information has been a topic of research in several communities over the past decade. The most common way to capture meeting information is through note-taking. However, fully writing down the content of a meeting is a difficult task, and can result in an inability to both take notes and participate in the meeting.

The existing tree-based mining method for discovering frequent patterns of human interactions in meeting discussions at same meeting[5][4]. Also existing algorithm applied only in the small tree dataset. The mining results, useful for analyzing and the comparison of meeting records. It can be used for understand about human interaction in meetings. The previous works not capture the meetings of the different categories. So we enhance our approach with various categories of tree dataset (i.e., for various categories of meetings).

III. HUMAN INTERACTIONS

The meaning of human interaction varies depending on the usage of the meetings or the types of the meetings. In this research, we mainly focus on the task-oriented interactions. The other communicative actions that concern the meeting and the group itself (e.g., when someone invited another participant to take the floor) are not included. We create a set of interaction types based on a standard utterance-unit tagging scheme. Like propose, comment, acknowledge, request information, ask opinion, positive opinion, and negative opinion. User proposes an idea with respect to a topic. For example, comment- a user comments to user a proposal, acknowledgement- acknowledge the proposal[1].

A. *Human Interaction Flow*

Human interaction flow is designed as the tree. An interaction flow is a list of all interactions in a discussion session with triggering relationship between them[1]. For this representation use labels for human interactions. L = {PRO, COM, ACK, REQ, ASK, POS, NEG}. Labels are abbreviated names of interactions, i.e., PRO-Proposing, COM-Commenting, ACK- Acknowledgement, REQ-Request Information, ASK-ask Opinion, POS- giving positive Opinion, and NEG- giving negative Opinion. The example human interaction tree shown in the fig. 2.

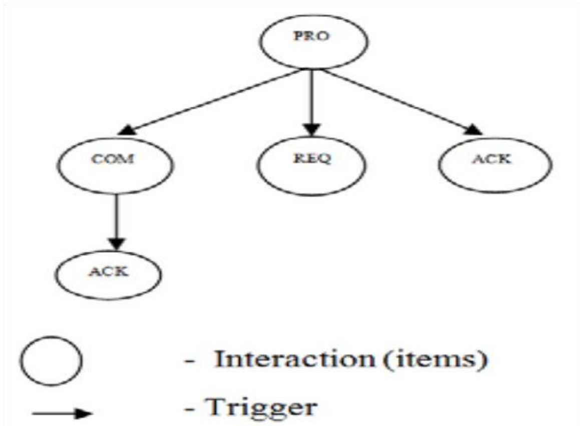

Fig. 2: Example Tree Representation for Human Interaction Flow.

IV. TREE BASED HUMAN INTERACTION MINING

A. Tree-Based Pattern Mining Algorithm

We designed an algorithm for interaction flow[1]. We used frequent tree pattern mining algorithm. For each tree in TD, the algorithm first exchanges the places of siblings (i.e., performs commutation processing) to generate the full set of isomorphic trees, ITD. The purpose of generating isomorphic trees is to ease string matching. The tree dataset is constructed by using Sting Encoding method.

1. String Encoding

To represent a Tree Dataset (TD) using Sting Encoding method[1]. The first node of the root is represented using "-", and subtree is represented by using "()", sibling relationships are represented using "*". The example sting code for Figure2 tree is represented by PRO-(COM-ACK)*REQ*ACK. Tree based mining algorithm works using Association rule mining under data mining. First, tree dataset (TD) is constructed. Then this algorithm calculates the support of each node (tree) in ITD. Then it selects the trees whose supports are larger than σ (min support). It finally outputs the frequent patterns as frequent trees.
Where,
TD - A dataset of interaction trees.
ITD - The full set of isomorphic trees to TD T - A Tree
t^k - A subtree with k nodes, i.e., K-subtree C^k - A set of candidates with k-nodes.
F^k - A set of frequent k-sub trees σ - A support threshold min sup
Algorithm1:
FITM (TD, σ) (Frequent Interaction Tree pattern Mining) Input: a tree database TD and a minsupport σ[7]
Output: all frequent tree patterns
Procedure:
Step1: In tree database TD, generate its full isomorphic trees dataset, ITD.
Step2: After database ITD generation, count the number of occurrences for each tree t.
Step3: Calculate the support of each tree.
Step4: Select the trees whose supports are larger than σ
(minsup).
Step5: Output the frequent trees.

B. Frequent Interaction Subtree Pattern Mining

This algorithm[1][7], extend tree based mining algorithm for subtree mining using the nature Apriori property. Because, through the subtree mining algorithm we can extract a more number of frequent trees as a parent-child subtrees. This algorithm works under the Apriori algorithm (candidate generation).

Basically this Apriori algorithm is a level-wise or an iterative algorithm. Find all 1-item frequent itemsets, then the 2-item frequent itemsets, which is continued until the last item. Here this mechanism, first calculates the support of each node and selects the nodes whose supports are larger than σ. To form the set of frequent nodes, F1 (Steps 2-3). It joins a frequent node to existing frequent i-subtrees to generate the set of candidates with i + 1 node (Steps 4-8). If there are any trees whose calculated supports are larger than σ , it selects them to form F^i +1 and repeats the procedure from Step 4. In Step 7, we join ti and t1 to generate the candidate subtree set of size.
Algorithm2:
FISTM (Frequent Interaction Sub Tree pattern Mining) Input: a tree database TD and a minsupport σ
Output: all frequent subtree patterns (as parent-child relationship)
Procedure:
Step1: i <-0
Step2: In tree database TD, calculate the support for each node
Step3: Select the nodes whose supports are larger than σ to form F1 (Candidate generation)
Step4: i<- i +1
Step5: For each tree t^i in F^i, do
Step6: For each node t^1 in F^1, do
Step7: Join the selected nodes ti and t1 to generate Candidates
Step8: Subtree Support Calculating (TD; t^i +1)
Step9: If there are any trees whose calculated supports are larger than the minsup, then select them to form F^i+1 and return to Step (4)
Step10: Else prune the other trees. And output the frequent subtrees whose supports are larger than σ

C. Various Categories of Datasets

In proposed system the tree mining algorithm is applied for extracting interaction pattern from various categories of meetings, debates, interviews and panel (different types of meetings). The common pattern from all types meetings and unique patterns of different types of meetings are analyzed. We investigate data mining techniques to detect and analyze frequent interaction patterns. We also develop several applications based on the discovered patterns. In this module develops various categories of the datasets in frequent common interaction mining by apply tree-based mining and subtree mining mechanisms in it.

D. Embedded Sub Tree Mining

This embedded sub tree mining is also similar to subtree mining. It also plans to discover embedded tree mining for hidden interaction pattern discovery. Embedded subtrees are a subtree, which allow not only parent child branches, but extract frequent subtrees in ancestor-descendant branches. It focuses on mining frequent embedded subtrees from databases of rooted labeled ordered sub trees[8].

1. Association Rule Mining

The association rule is a popular method for discovering interesting relationships among items in large databases. In our approach the association rules mining are used to discovering the confidence and support value. Support value is calculated between the number of occurrences of tree or sub-tree and the total number of the trees in the dataset of interaction trees. The support value is calculated using the following formula

Support = (Number of occurences of required tree / Total number of all trees in tree dataset)
Frequent Value = (Maximum no. of comments present in data set/ total no .of comments)

The value is displayed after the embedded sub tree mining is performed. Like subtree candidate generation process. But here the candidates are joined from the ancestor and descendent level nodes of the subtree mining. So the performance of frequent pattern extraction is improved. But the middle node is hidden[6]. Because, the middle node is not a frequent node in a subtree. Only the ancestor descendent nodes are frequent in this algorithm.

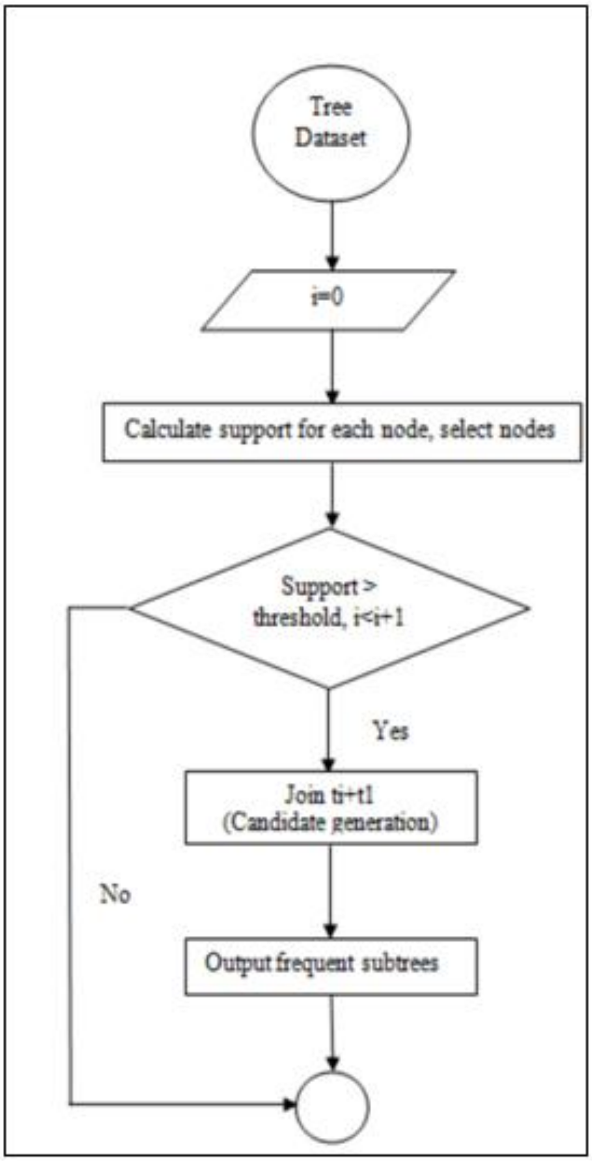

Fig. 3: Flow chart 1: Frequent interaction subtree pattern mining

E. Temporal Mining

Temporal data mining techniques used to extract the temporal patterns (i.e., particular time period trees). From captured content of human interaction in meetings. That is extracting patterns in a particular time period from each and every participant's interaction[7]. The human integration activities varied based on time and experience of events. To extract temporal pattern mining used hidden markov model along with tree mining algorithm. HMM allow to estimate probabilities of observed events in pattern mining. In our approach we estimate the transition and emission probabilities between the observed events. Two probabilities are used in the system. Transitional probabilities are used to find the probabilities for movement from one state to another state. That is unobserved events in the system, observation probability are used to find the observed events in the system called emission probability. HMM is a Markov model for which have a series of observed output and the probability of generating an output observation as a function of our hidden state or indirect information. So based on the probability calculation can extract individual people or participants particular information efficiently

V. CONCLUSION

Tree-based mining method for discovering frequent patterns of human interaction in meeting discussions. The mining results useful for analyzing and comparison of meeting records. It is also plan to explore embedded tree mining for hidden interaction pattern discovery. It is valuable to capture various categories of meetings for analysis such as panel, debate, and interview. Because the current meetings are all task oriented. The goal is to discover frequent interaction trees, and the behavior of the algorithms on the data set, using the behavior threshold.

The proposed work develops several applications based on the discovered patterns. Embedded subtrees are subtrees, which allow ancestor-descendant branches. Hidden markov model is used to extract the temporal pattern in case of the time period along with the tree mining and finally plan to incorporate more meeting content in both amount and category. Frequent interaction tree mining (FITM), Frequent interaction subtree mining (FISTM), embedded subtree(EST), Hidden Markov Model (HMM). The number of frequent pattern and minimum support value is the most considerable[7]. Performance of extracting the frequent patterns based on the time periods and its probabilities of HMM is faster other then three models. The human interaction activities varied based on the time to find the frequent patterns.

REFERENCES

[1] Christian Becker, Zhiwen Yu,"*Tree-Based Mining for Discovering Patterns of Human Interaction in Meetings*", Proc. Eighth IEEE Int’l Conf. Knowledge Discovery and Data Mining (PAKDD’10), pp. 107-115,2012.

[2] Z.Yu, Y. Nakamura,“*Smart Meeting Systems: A Survey of State-of-the-Art and Open Issues*”, ACM Computing Surveys, Vol. 42, No. 2, article 8, Feb. 2010

[3] S.Junuzovic, R. Hegde, Z. Zhang, P. Chou, Z. Liu, Zhang,“*Requirements and Recommendations for an Enhanced Meeting Viewing Experience*”, Proc. ACM Int’l Conf. Multimedia, pp -539- 548,2008

[4] K. Otsuka, H. Sawada, J. Yamato,“*Automatic Inference of Cross-Modal Nonverbal Interactions in Multiparty Conversations*”, Proc. Int’l Conf. Multimodal Interfaces (ICMI ’07), pp. 255- 262, 2007.

[5] I. Mccowan, D. Gatica-Perez, S. Bengio, G. Lathoud, M. Barnard, D. Zhang, “*Automatic Analysis of Multimodal Group Actions in Meetings*”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 27, No. 3, pp. 305-317, Mar.2005.

[6] L. Anolli, S. Duncan Jr, M.S. Magnusson, G. Riva, “*The Hidden Structure of Interaction: From Neurons to Culture Patterns*”, Emerging Communication: Studies in New Technologies and Practices in Communication, IOS Press, Apr. 2005.

[7] 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.

[8] C. Wang, M. Hong, J. Pei, H. Zhou, W. Wang, 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.