

# Graph Mining – To Extract Information Using Human Interaction based on Pattern Discovery

Dr. D. Durga Bhavani

Professor, Department of Computer Science and Engineering, NMREC, Hyderabad, India

**Abstract:** *An interaction tree pattern mining algorithms was projected to analyze tree structures and extract interaction flow patterns. In this paper, the tree based mining for human interaction flow in a discussion session is represented as a tree. Tree-based interaction mining algorithms are designed to analyze the structures of the trees and to extract interaction flow patterns. This paper deals with a mining algorithm to extract frequent patterns of human interaction explore tree mining for hidden interaction pattern discovery using classification of mining algorithms from captured content of time series of many meetings in particular time periods years ago. The paper also provides the efficient method to extract the information priority wise algorithms compare to earlier technique. The work focuses on discovering higher level knowledge about human interaction. This paper also deals to extract several interesting patterns that are useful for the interpretation of human behavior in meeting discussions, such as determining frequent interactions, typical interaction flows, and relationships between different types of interactions.*

**Keywords:** Human Interaction, Partitional Clustering, Pattern Discovery, Mining Algorithms, Tree structure Graph

## 1. Introduction

This paper deals 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. 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.

Mining scores based on historically analysis of likelihood of fraud were custom developed based on text entered in the claims reports and details based on the claims. Reports are generated on a web based application layer fed into the enterprise resource planning application which is used by the client and commonly found in most of the larger insurance companies. Many text mining applications give users open ended freedom to explore text for meaning. Text mining can be used as a deeper more penetrative method which goes beyond escalations of possible search interests and to sense the mood of the written text. Text mining is the method that supports users to find useful information from a large amount of a digital text data should retrieve the information that users require with relevant efficiency. Information Retrieval has the objective of automatically retrieving as many relevant documents as possible filtering out irrelevant documents at the same time. However Information retrieval based system do not adequately provide users with what really need. Many

text mining methods have been developed in order to achieve the goal of retrieving for information users, process of extracting discovery pattern consist of following Data Selection, Data Processing, Data Transaction, Pattern Discovery & Pattern Evaluation. The ability to search for keywords in a collection is widespread such search only slightly supports discovery because the user has to decide and can suggest interesting patterns to look at and the user can then accept or reject these pattern as interesting. Data mining, which is a powerful method of discovering new knowledge, has been widely adopted in many fields, such as bioinformatics, marketing, and security [1][4].

In this study, we investigate data mining techniques to detect and analyze frequent interaction patterns; we hope to discover various types of new knowledge on interactions. Human interaction flow in a discussion session is represented as a tree. Inspired by tree-based mining, we designed interaction tree pattern mining algorithms to analyze tree structures and extract interaction flow patterns. An interaction flow that appears frequently reveals relationships between different types of interactions. Mining human interactions is important for accessing and understanding meeting content. First, the mining results can be used for indexing meeting semantics, also existing meeting capture systems could use this technique as a smarter indexing tool to search and access particular semantics of the meetings. Second, the extracted patterns are useful for interpreting human interaction in meetings. Cognitive science researchers could use them as domain knowledge for further analysis of human interaction. Moreover, the discovered 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.

## 2. Related Work

Information retrieval system identify the documents in a collection which match a user query, search engine which allows identification of a set of documents that relate to a set

of key words. Information extraction is the process of automatically obtaining structured data from an unstructured natural language document. Term analysis identifies the terms in a document where a term may consist of one or more words especially useful for documents that contain many complex multi word terms such as scientific. Named-entity recognition identifies the names in a document such as the names of people systems are also able to recognize dates and expressions of time quantities and associated unit percentages and fact extraction which identifies and extracts complex facts from documents such facts could be relationship between entities or events.

Data mining is the process of identifying patterns in large sets of data when used in text mining is applied to the facts generated by the information extraction phase and the result of data mining process are put into another database that can be queried by the end-user via a suitable graphical interface network of protein interactions. Electronic discovery refers to discovery deals with the exchange of information in electronic format and agreed upon processes and often reviewed for privilege and relevance before being turned over to opposing.

Data are identified as relevant and extracted using digital forensic procedures and is reviewed using a document review and useful for its ability to aggregate. Electronic information is different from paper information because of its intangible form usually accompanied by metadata that is not found and can play an important part as evidence. Providing a text mining for science requires a new means of collaboration between existing and future stakeholders to accept data and text mining as being effective and acceptable processes such mining does not eliminate any significant role currently being performed by stakeholders that it does not raise challenges and barriers to text mining applications that it does not threaten publishers [2].

Text mining is believed to have a considerable commercial value particularly true in scientific disciplines in which highly relevant information is often contained within written text. A distributed model raises issues around data normalization of performance levels of other standardization issues requires conformity by all involved to common metadata standards to allow effective cross reference and indexing. In support of text mining one can see the emergence of the cloud as a mechanism for processing large amounts of data using the existing powerful computer resources made available by organizations such as Amazon, Microsoft, HP, etc.

### 3. Implementation of Algorithms

Purpose of the analysis in some instances the extraction of semantic dimensions alone can be useful outcome if it clarifies the underlying structure of input documents. To reiterate text mining can be summarized as a process of numeric text.

#### 3.1 Tree pattern mining algorithms

An interaction flow that appears frequently reveals relationships between different types of interactions. Mining human interactions is important for accessing and understanding meeting content. To analyze tree structures and extract interaction flow patterns. A tree-based mining method is used for discovering frequent patterns of human interaction in meeting discussions. The mining results would be useful for summarization, indexing, and comparison of meeting records. They also can be used for interpretation of human interaction in meetings.

#### 3.2 Algorithms for Pattern Discovery

With the representation model and annotated interaction Flows, we generate a tree for each interaction flow and thus build a tree data set. For the purpose of pattern discovery, we first provide the definitions of a pattern and support for determining patterns. In developing our frequent sub tree discovery algorithm, we decided to follow the structure of the algorithm for pattern discovery used for finding frequent item sets, because it achieves the most effective pruning compared with other algorithms.

#### 3.3 Admin Analysis

The Main details of the users can not viewed by users, that type of process is maintained by the admin process. Admin process is the process that maintains the process and users details. Admin process can view the process as tree based structure. So the admin can easily identified by the human interactions. Human interaction process can be viewed by admin by the following structure based elements are various elements which are described in the following modules. Session tree process is the process that is used to avoid the repeated data in session database. And the process provides the tree based structure. So the admin identified the main problem in the particular topic. All process such as PC purchase, Trip planning, Soccer and job can be viewed by the admin process. Graph is another process for the admin view. This process is also related to the session tree concept. But is process only provides the separate graph view. So the admin can easily maintain the process.

#### 3.4 Interaction Flow Construction

Interaction flow construction create an environment based on the interaction defined and recognized, we now describe the notion of interaction flow and its construction. An interaction flow is a list of all interactions in a discussion session with triggering relationship between them. We create an application based on it. In the application we have authentication process. For authentication process we build Login process, which is used for enter the process and register the new users. This process is produced for both users and admin process. All users details can be stored in the database elements. So, unwanted users cannot easily access this Login process. Homepage is used for the login. Registration process requires the Name, Details, address, phone number and email id

## 4. Analysis of Problem

Tree based mining method for discovering frequent patterns of human interactions in meeting discussion mining technique analyze the comparison of meeting records and used to understand about human interaction in meetings. Discovering a pattern semantic knowledge is significant for understanding and interpreting how the user interact in a meetings like business, commercial & Academic purpose. Common way of capturing the information is through note taking however manually written down the content of a meeting is a difficult task and can result in an ability to both take note and participate in the meeting.

### 4.1 Human Interaction

Human interaction varies depending on the usage of the meetings or the types of the meetings and task oriented interactions communicative actions that concern the meeting and the group itself. Set of interaction types based on a standard unit scheme comment acknowledge request information ask opinion positive opinion and negative opinion. User proposes an idea with respect to a subject or proposal. Representation use labels for human interactions are abbreviated names of interactions that commenting request information ask opinion giving positive giving negative opinion.

### 4.2 Tree Based Mining Technique

Finding frequent item sets in data warehouse operation of association rule mining, frequent patterns have many useful applications in markup language marketing banking networking routing. A mining method to extract frequent items of human interaction based mining on the extracted content of interactions. Human interactions such as proposal or giving comments opinions are constructed as a priority like a tree, tree based interaction mining algorithm are designed to analyze the structures of the trees and to extract frequent patterns in a tree dataset. Capturing all of informal meeting information is omitted by using tree based mining approach to extract frequent patterns of human interactions based on the captured content is of human participated meetings, mining results can be used for context purpose meeting semantics also existing meeting capture systems use this technique as a smarter indexing tool to search and access particular semantics of the meetings.

## 5. Analysis of Mining Algorithms

Supervised learning is to use the available data to build one particular variable of interest in terms of rest of data. A number of classification algorithms can be used for anomaly detection. Proposes the use of ID3 Decision tree classifiers to learn a model that distinguishes the behavior of intruder from the normal user's behavior. Supervised (Classification): Decision Tree, Bayesian classification, Bayesian belief networks, neural networks etc. are used in data mining based applications [5].

Unsupervised learning is where no variable is declared as target the goal is to establish some relationship among all variables. Unsupervised learning studies how systems can learn to represent particular input patterns in a way that reflects the statistical structure of the overall collection of input patterns. The unsupervised learner brings to bear prior biases as to what aspects of the structure of the input should be captured in the output. In this paper combination of applications Supervised & Unsupervised has been combined together used to solve the problem of Network Anomaly Data. A Very rare case both the Techniques have been combined. Unsupervised (Clustering): Association Rules, Pattern recognition, Clustering Techniques. Here, clustering Technique is one of the media to Network Anomaly data [3][4].

### 5.1 Classification Techniques

In Classification, training examples are used to learn a model that can classify the data samples into known classes. The Classification process involves following steps:

- a) Create training data set.
- b) Identify classes and attributes
- c) Find useful attributes for classification (Relevance ranking analysis).
- d) Acquire a model using training data set.
- e) Use the model to classify the unknown data samples.

### 5.2. Clustering Technique

Cluster is a number of similar objects grouped together. It can also be defined as the organization of dataset into homogeneous and/or well separated groups with respect to distance or equivalently similarity measure. Cluster is an aggregation of points in test space such that the distance between any two points in cluster is less than the distance between any two points in the cluster and any point not in it. There are two types of attributes associated with clustering, numerical and categorical attributes. Numerical attributes are associated with ordered values such as height of a person and speed of a train. Categorical attributes are those with unordered values such as kind of a food items and brand of cloths.

Clustering is available in flavors of

- Hierarchical
- Partition

In hierarchical clustering the data are not partitioned into a particular cluster in a single step. Instead, a series of partitions takes place, which may run from a single cluster containing all objects to n clusters each containing a single object. Hierarchical Clustering is subdivided into agglomerative methods, which proceed by series of fusions of the n objects into groups, and *divisive* methods, which separate n objects successively into finer groupings[3][5].

For the partitional can be of K-means & K-mediod. The purpose solution is based on K-means clustering combine with Id3 Decision Tree type of Classification under mentioned section describes in details of K-means & Decision Tree. K-means is a centroid based technique, each

cluster is represented by the center of gravity of the cluster so that the intra cluster similarity is high and inter cluster similarity is low. This technique is scalable and efficient in processing large data sets because the computational complexity is  $O(nkt)$  where  $n$ -total number of objects,  $k$  is number of clusters,  $t$  is number of iterations and  $k \ll n$  and  $t \ll n$ .

### 5.3. K-mean algorithm

K-mean is the most popular partitioning method of clustering. It was firstly proposed by MacQueen in 1967. K-mean is an unsupervised, non-deterministic, numerical, iterative method of clustering. In k-mean, each cluster is represented by the mean value of objects in the cluster. Here we partition a set of  $n$  object into  $k$  cluster so that intercluster similarity is low and intracluster similarity is high. Similarity is measured in term of mean value of objects in a cluster.

- 1) Select  $k$  centroids arbitrarily for each cluster  $C_i, i \in [1, k]$
- 2) Assign each data point to the cluster whose centroid is closest to the data point.
- 3) Calculate the centroid  $C_i$  of cluster  $C_i, i \in [1, k]$
- 4) Repeat steps 2 and 3 until no points change between clusters.

Main drawback of K means is that one must specify the clusters in advance and further the algorithm is very sensitive of noise, mixed pixels and outliers. The definition of means limit the application to only numerical variables. We choose k-means because it is data driven with relatively few assumptions on the distributions of underlying data.

### 5.4. Decision Tree

Decision tree support tool that uses tree-like graph or models of decisions and their consequences, including event outcomes, resource costs, and utility. Commonly used in operations research, in decision analysis help to identify a strategy most likely to reach a goal. In data mining and machine learning, decision tree is a predictive model that is mapping from observations about an item to conclusions about its target value. The machine learning technique for inducing a decision tree from data is called decision tree learning.

Assume a tree is classified into five leaf nodes. In a decision tree, each leaf node represents a rule. The following rules are applied.

Rule 1: If it is sunny and the humidity is high then do not play.

Rule2 : If it is sunny and the humidity is normal then play.

Rule3: If it is overcast, then play.

Rule4: If it is rainy and wind is strong then do not play.

Rule5: If it is rainy and wind is weak then play.

### 5.5 ID3 Decision Tree

Iterative Dichotomiser is an algorithm to generate a decision tree invented by Ross Quinlan, based on Occam's razor. It prefers smaller decision trees (simpler theories) over larger ones. However it does not always produce smallest tree, and

therefore heuristic. The decision tree is used by the concept of Information Entropy

### 5.6. Association Mining

Finding all relevant occurrence relationship is called association, it is market basket data analysis to discover items purchased by customers in a market are associated. Association mining algorithm developed with different mining efficiencies, which find the same set of rules though their computational efficiencies and memory requirements may be different in two steps. Frequent item is an item set that has transaction support, confident association rule is with confidence. Association mining is the threshold used to prune the search space and to limit the number of frequent item set and rules generated [1]. But using only a single implicitly assumes that all items in the data are of the same nature similar frequencies in the database. In some other applications items appear very frequently in the data that perform if the minsup is set too high not find rules that involve infrequent items or items are rare in the data, to find rules that involve both frequent and rare items have to set the minsup very low.

$L_k$ : set of frequent item set of size  $k$  with min support

$C_k$ : set of candidate item set of size  $k$  potentially frequent itemset

$L_1$  is frequent items for  $(k=1; L_k \neq \emptyset; K++)$  do begin s

## 6. Conclusion

Identification of negative points in topic is very tough and increases the repeated data. The work focuses on discovering higher level knowledge about human interaction. In our proposed system T-pattern technique is used to discover hidden time patterns in human behavior. To conduct analysis on human interaction in meetings and address the problem of discovering interaction patterns from the perspective of data mining. It extracts simultaneously occurring patterns of primitive actions such as gaze and speech. Mining tree method for discovering frequent patterns of human interaction in communication of business requirement, analysis shows the comparison of existing system mining algorithms used to extract meeting records. It is explore tree mining for hidden interaction pattern discovery.

## 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] J. Han, M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, March 2006.
- [4] A. K. Jain and R. C. Dubes. Algorithms for Clustering Dasta. Printice Hall, 1988.
- [5] A. K. Jain, M. Narasimha Murty, and P.J. Flynn. Data Clustering: A Review. ACM Computing Surveys, 31(3):264-323, 1999.