

# COMPARATIVE ANALYSIS OF HUMAN INTERACTION PATTERN MINING APPROACHES

S. Uma<sup>1</sup> and J. Suguna<sup>2</sup>

<sup>1</sup>Department of Information Technology, C.M.S. College of Science and Commerce, India

<sup>2</sup>Department of Computer Science, Vellalar College for Women, India

## Abstract

*Opinion Mining and Sentiment Analysis in Natural Language Processing (NLP) are challenging, as they require deep understanding. Understanding involves methods that could differentiate between the facts of explicit and implicit, regular and irregular, syntactical and semantic language rules. Researches oriented towards Natural Language Processing and Sentiment Analysis have many unresolved problems like co-reference resolution, negation handling, anaphora resolution, named-entity recognition, and word-sense disambiguation. This paper is proposed to develop an Optimized Partial Ancestral Graph (O-PAG) which is capable of mining patterns in human interactions and compare it with an existing tree based pattern mining approach. The experimental results are exposed to number of frequent interactions made and execution time. Results indicate that the overall performance can reach considerable improvements on using O-PAG approach.*

## Keywords:

*Frequent Pattern Mining, Sequential Pattern, Enhanced PCA, Enhanced ABC, Pattern Mining*

## 1. INTRODUCTION

Frequent pattern mining mines more complex forms of frequent patterns. It considers user preferences or constraints to speed up the mining process. It has reached far beyond the basics due to substantial research, numerous extensions of the problem scope, and broad application studies. An in-depth coverage of methods for mining many kinds of patterns is included elaborating on: multilevel patterns, multidimensional patterns, patterns in continuous data, rare patterns, negative patterns, constrained frequent patterns, frequent patterns in high-dimensional data, colossal patterns, and compressed and approximate patterns. Other pattern mining themes, including mining sequential and structured patterns and mining patterns from spatiotemporal, multimedia, and stream data, are considered more advanced.

Pattern mining is a more general term than frequent pattern mining since the former covers rare and negative patterns as well. But, when there is no ambiguity, the two terms are used interchangeably. In addition to mining for basic frequent itemsets and associations, advanced forms of patterns can be mined such as multilevel associations and multidimensional associations, quantitative association rules, rare patterns, and negative patterns. Users can also mine high-dimensional patterns and compressed or approximate patterns. Frequent pattern mining has many diverse applications, ranging from pattern-based data cleaning to pattern-based classification, clustering, and outlier or exception analysis [1].

Mining human Interaction in meetings can be used to determine the reactions of persons in different scenarios [2]. The four major tasks involved in sentiment analysis are identifying

opinions, features extraction, sentiment classification and summarization of results [3].

The first task analyzes the statement and identifies whether the statement contains opinion or not. It extracts only the statement containing opinions for further processing. Feature extraction identifies feature or entity that can be a place, object, person or anything. Next step is the process of classification and it determines the type of opinion. The final task is the report to the user (Fig.1).

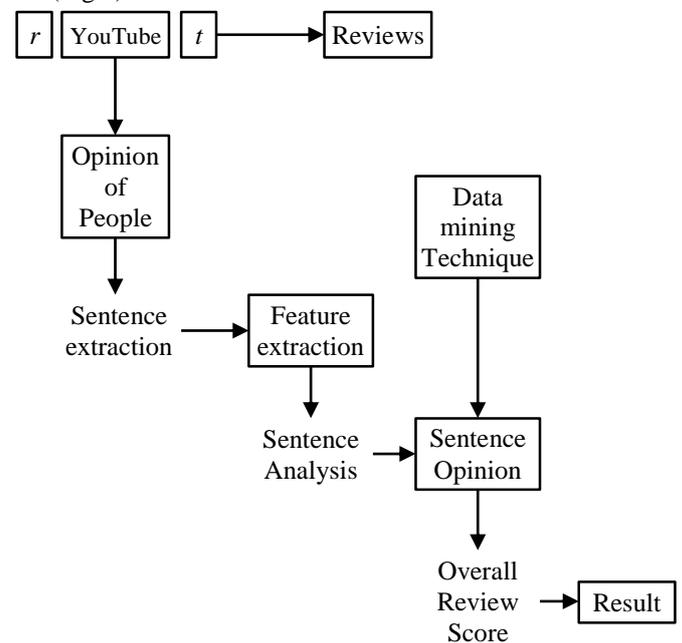


Fig.1. Process of Sentiment Analysis

## 2. LITERATURE SURVEY

Kouloumpis et al. [4] made an examination on linguistic features utility for recognizing the sentiments of twitter messages. The creator has assessed whether the data gathered from hashtags and emoji are valuable for preparing twitter's sentiment classifiers and presumed that the emotion preparing data advantages are reduced when small scale blogging features are utilized [4].

They additionally proposed that on ascend to long range interpersonal communication age, has been a gigantic substance of client produced report. A large number of individuals share their view and musings in Micro blogging destinations ordinary on account of its short and basic method for expression. Twitter, face book, MySpace, YouTube are the best models, where individuals post their continuous experience and opinions in every one of the subjects which have turned out to be one of the most significant uses of Web 2.0 that enable individuals to assemble association with other individuals through system in a simple

manner and enable them to share different sorts of data like picture sharing, sites and so forth.

Fillmore et al. [5] portrayed the data structure and standards of FrameNet for speaking to lexical meaning of frame based analysis. The lexical liking is marginally much superior to anything the watchword spotting approach where than considering influence words; it relegates arbitrary words a probabilistic proclivity for a specific emotion. The task of these probabilities depends on the linguistic [5]. So somewhat utilizing this approach sentiment could be investigated.

Aggarwal et al. [6] made a study on sentimental analysis and opinion mining by contemplating similar analysis on different data mining calculations and Machine learning strategies. The creator talked about that feature extraction assumes a significant job in Opinion mining and Sentiment analysis. A significant assignment of opinion mining is to extricate opinion of the individuals on features of a substance. Feature extraction is confronting various issues and feature extraction is the most troublesome assignment in sentimental analysis [6].

Yang et al. [7] presented that fuzzy clustering theory and source database was built for micro-blog. The creator utilized HowNet, a fundamental emotion set to ascertain loads of micro-blog emotion words and emotional lexicon was fabricated. At long last, utilizing fuzzy clustering calculation, assessment esteems were got from clients in various timeframes and arranged powerfully utilizing insights. Best grouping outcome was acquired to the various needs and take the fitting measures. As per creator Human interaction is the most significant qualities of social elements in gathering meetings. It is only a discourse that characterizes open activities by the members of the gathering comparing to the present topic. Mining human interactions is significant for understanding gathering content [7].

Human interaction meeting are significant action of correspondence where new thoughts alert, issues talked about and options found. Human Interaction can be caught by directing social gathering and can perceive the sort of interaction with their answer like giving a positive and negative sentiment, proposing a few thoughts, giving remarks and so forth.

### 3. HUMAN INTERACTION PATTERN MINING APPROACHES

#### 3.1 TREE BASED APPROACH

Meeting is an important factor in the development of business and it can be successful when it is well organized. To find out if the meeting is well organized or not is possible by analyzing human behavior or human interaction in the meeting. The formation of well-defined dictionary for the relevant meetings is the challenging problem and it is formed based on the people interaction [8].

Mining of human interaction during the conduct of a meeting gives a chest of resources that gives detailed information regarding the meeting and how things are perceived by the participant regarding the meeting. These responses can help management in decision making and quality improvement. The tree based mining algorithm is used to review the structure of the parent child relationship in tree. The different types of interactions that happen in a meeting are:

- *Propose*: Novel ideas or opinions that are shared by participant.
- *AskOpinion*: The opinions are asked from the participants for effective meeting.
- *PosOpinion*: Sharing the positive attitude or opinion for the proposed idea by participant.
- *NegOpinion*: Sharing the negative attitude or opinion for the proposed idea by participant.
- *Acknowledgement*: The participant agrees or disagree the proposal or idea.
- *Comment*: Commenting about others interactions or behavior.
- *RequestInformation*: The participant requesting the information regarding the idea or issue.

The classification in patterns discovery are based on the support vector machine, Bayesian Net, Bayes, and Decision Tree. The parameters used to construct the interaction flow include a session which is a unit of a meeting that initiates with a natural interaction and concludes with an interaction that is not followed by any reactive interactions. In session, the interactions are normally classified as a spontaneous interaction and reactive interaction.

#### 3.2 FREQUENT PATTERN MINING (FPM)

FPM is a tree based mining method which will discover frequent patterns of human interaction from meeting discussions [8]. For each tree in tree dataset, isomorphic trees are generated with the number of occurrences. The support of each tree will be calculated and the trees whose supports are larger than the threshold value are selected. The isomorphic trees within them are detected. If many trees are isomorphic, one of them is selected and the others are discarded and the frequent trees are given as output. FPM will generate more frequent patterns but will miss some important frequent pattern as it does not capture all triggering relations.

#### 3.3 OPTIMIZED PARTIAL ANCESTRAL GRAPH (O-PAG) APPROACH

##### 3.3.1 Enhanced Principal Component Analysis (EPCA):

Principal Component Analysis (PCA) is a dimension-reduction tool that can be used to reduce a large set of variables to a small set that still contains most of the information in the large set. Principal Component Analysis (PCA) is a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components [9].

In human interactions, patterns can be derived from multiple interactions. Such interaction dataset will have confusing and irrelevant words in sentences. EPCA will reduce such complex data to a lower dimension and eliminate the variations present in the data [10].

##### 3.3.2 Enhanced Artificial Bee Colony Approach (EABC):

ABC algorithm is a new swarm intelligent algorithm, which proposed by Karabog in Erciyes University of Turkey in 2005. Since ABC algorithm is simple in concept, easy to implement, and has fewer control parameters, it has been widely used in many

optimization applications such as protein tertiary structures, digital IIR filters, artificial neural networks and others. The minimal model of foraging selection that leads to the emergence of collective intelligence of honey bee swarms consists of three essential components: food sources, employed foragers and unemployed foragers [11].

There are two basic behaviours based on recruitment to a food source and the abandonment of a food source.

- 1) *Food Sources*: it represents a position of solution of optimization problem, the profitability of food source are expressed as fitness of the solution.
- 2) *Unemployed Foragers*: there are two types of them, scouts and onlookers. Their main task is exploring and exploiting food source. At the beginning, there are two choices for the unemployed foragers:
  - i. It becomes a scout - randomly search new food sources around the nest;
  - ii. It becomes an onlooker - determine the nectar amount of food source after watching the waggle dances of employed bee, and select food source according to profitability.
- 3) *Employed Foragers*: the honeybees found food source, which also known as the employed bees, are equal to the number of food sources. The employed bees store the food source information and share with others according to a certain probability. The employed bee will become a scout when food source has been exhausted.

In most of the unconstrained ABC algorithms the role of population initialization is ignored. In order to have powerful algorithm preliminary solutions must be generated uniformly within the search space. The uniformly distributed initial solutions help to produce at least some points in the neighborhood of global solution. To obtain that, Partial Least Square (PLS) dimension reduction mechanism is used. PLS is based on correlation. PLS is very popular in areas like chemical engineering, where predictive variables often consist of many different measurements in an experiment and the relationships between these variables are ill-understood [12].

In main loop after producing a new solution in employee bee phase, Deb's constrained handling method is adopted in order to adapt ABC algorithm for solving constrained optimization problems instead of using greedy selection in unconstrained ABC [13].

**Algorithm 1: Enhanced ABC algorithm**

- Step 1:** Initialize the population using PLS
- Step 2:** Evaluate the initial population
- Step 3:** Cycle=1
- Step 4:** Repeat
- Step 5:** Employed bee phase
- Step 6:** Apply Deb's mechanism to select between and
- Step 7:** Calculate the probability values for the solution
- Step 8:** Onlooker bee phase
- Step 9:** Apply Deb's mechanism
- Step 10:** Scout bee phase
- Step 11:** Memorize the best solution achieved so far

**Step 12:** Cycle = cycle + 1

**Step 13:** Until cycle = maximum cycle number

**3.4 PARTIAL ANCESTRAL GRAPH (PAG) MEET**

Partial Ancestral Graph (PAG) meet will find temporal frequent interactions. It will generate a set of all frequent nodes and then expand these nodes with new root, new level, new node and new edge. The sub-graphs containing siblings may not be connected without the presence of their common ancestor in a tree. So, if a common ancestor is not frequent, tree based mining method will fail to mine them as frequent pattern. Also it will not distinguish multiple interactions but PAG meet will identify such multiple interactions and extract both temporal and triggering relations [14].

**3.5 OPTIMIZED PARTIAL ANCESTRAL GRAPH (O-PAG)**

The proposed approach performs dimensionality reduction, optimization and generates all frequent patterns capturing both direct parent-child relations as well as ancestral temporal relations. It is termed as Optimized Partial Ancestral Graph (O-PAG).

Each phase in O-PAG depends on its previous phase to derive useful and valuable patterns in human interaction in a meeting. The initial phase in designing an O-PAG comprises of preprocessing using optimization techniques. The datasets that is chosen for the study have certain amount of irrelevant data that needs to be mined and hence as a first step EPCA mechanism is used to remove such irrelevant data. When EPCA completes the reduction of deviated discussion contents, EABC algorithm is applied.

The next phase is ancestral graph generation. The above preprocessed data is now fed into PAG meet and interaction flow in meetings are generated.

**4. RESULTS AND DISCUSSIONS**

Dataset is gathered from meetings that are conveyed through on-line forums and documented discussions from various data sources. The Table.1 shows the categorization of datasets.

Table.1. Dataset Information

Document Type	Number of documents	Maximum number of conversations per document
Mini dataset	2000	8
Large dataset	2000	250
Very large dataset	2000	5000

The datasets are focused on the topics namely corruption, democracy, defense, expenditure, social expenditure, backwardness, women quota bill, communal and religious reservations, retirement age, and elections etc. First observation was done by using FPM, which is a tree based approach. Another observation was done by using O-PAG approach.

Detecting more number of interactions gains advantage towards discovering frequent pattern mining. From Table.2, it is shown that the number of discovered frequent subtrees is increased in O-PAG approach than the tree based approach and the same is depicted in Fig.2.

Table.2. Support Threshold vs. Number of Discovered Frequent Subtrees

Support Threshold	Tree Based Approach	Optimized Partial Ancestral Graph
1	8200	14172
2	3800	5525
3	2300	6085
4	1902	5551
5	1704	5361
6	1203	5145
7	1101	4491
8	902	4662
9	753	4448
10	401	3793

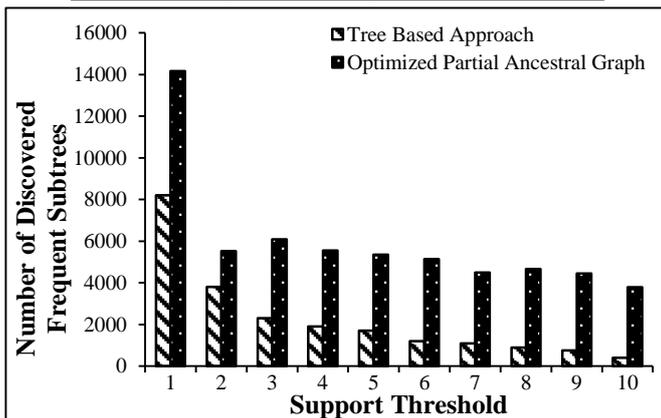


Fig.2. Support Threshold vs. Number of Discovered Frequent Subtrees

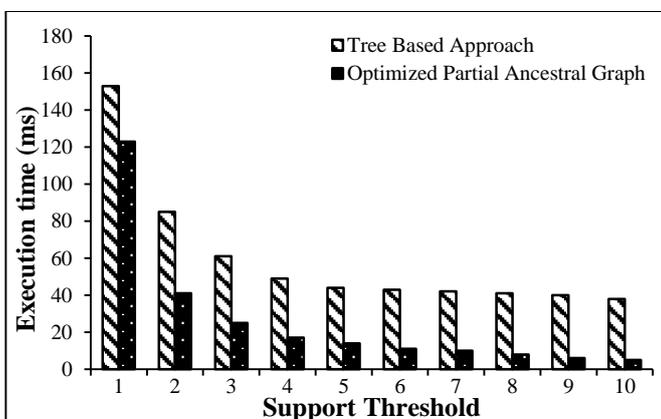


Fig.3. Support Threshold vs. Execution Time

The Table.3 illustrates execution time of tree based approach and O-PAG approach. It can be clearly understood that the O-

PAG approach meet consumes less execution time than tree based approach and it is noteworthy fact that when support threshold value increases the execution time is decreased and the same is flashed in Fig.3.

Table.3. Support Threshold vs. Execution Time

Support Threshold	Tree Based Approach	Optimized Partial Ancestral Graph
1	153	123
2	85	41
3	61	25
4	49	17
5	44	14
6	43	11
7	42	10
8	41	8
9	40	6
10	38	5

### 5. CONCLUSION

This paper is proposed to apply EPCA and EABC Algorithms to remove irrelevant data in meeting discussions and construct an Optimized Partial Ancestral Graph (O-PAG) and compare it with existing tree based approach that is capable of mining patterns in human interactions. The results indicate that the proposed O-PAG approach outperforms tree based approach in terms of execution time and number of discovered frequent sub-trees.

### REFERENCES

- [1] Z.W. Yu, Z.Y. Yu, H. Aoyama, M. Ozeki and Y. Nakamura, "Capture, Recognition, and Visualization of Human Semantic Interactions in Meetings", *Proceedings of 8<sup>th</sup> IEEE International Conference on Pervasive Computing and Communications*, pp. 107-115, 2010.
- [2] S. Uma and J. Suguna "Tree-Based Weighted Interesting Pattern Mining approach for Human Interaction Pattern Discovery", *International Review on Computers and Softwares*, Vol. 8, No. 11, pp. 2570-2575, 2013.
- [3] Henrique Siqueira and Flavia Barros, "A Feature Extraction Process for Sentiment Analysis of Opinions on Services", *Proceedings of International Workshop on Web and Text Intelligence*, pp. 1-10, 2010.
- [4] Efthymios Kouloumpis, Theresa Wilson and Johanna D. Moore, "Twitter Sentiment Analysis: The Good the Bad and the OMG!", *Proceedings of 5<sup>th</sup> International Conference on Weblogs and Social Media*, pp. 538-541, 2011.
- [5] Charles J. Fillmore and Collin Baker, "A Frames Approach to Semantic Analysis", Oxford University Press, 2018.
- [6] Charu C Aggarwal and Cheng Xiang Zhai, "Mining Text Data", Springer, 2012.
- [7] Li Yang, Xinyu Geng and Haode Liao, "A Web Sentiment Analysis Method on Fuzzy Clustering for Mobile Social

- Media Users”, *Journal on Wireless Communications and Networking*, Vol. 128, pp. 1-16, 2016.
- [8] Christian Becker and Zhiwen Yu, “Tree-Based Mining for Discovering Patterns of Human Interaction in Meetings”, *Proceedings of IEEE International of Conf. Knowledge Discovery and Data Mining*, pp. 107-115, 2012.
- [9] David G. Kleinbaum, L.L. Kupper, A. Nizam and K.E. Muller, “*Applied Regression Analysis and Multivariable Methods*”, Duxbury Press, 2007.
- [10] S. Uma. And J. Suguna, “Human Interaction Pattern Mining using Enhanced Principal Component Analysis”, *International Journal of Informative and Futuristic Research*, Vol. 2, No. 7, pp. 2279-2289, 2015.
- [11] W. Gao, S. Liu and L. Huang, “A Global best Artificial Bee Colony Algorithm for Global Optimization”, *Journal of Computational and Applied Mathematics*, Vol. 236, No. 11, pp. 2741-2753, 2012.
- [12] Dervis Karaboga and Bahriye Basturk, “Artificial Bee Colony Optimization Algorithm for Solving Constrained Optimization Problems”, *Proceedings of International Fuzzy Systems Associations*, pp. 789-798, 2007.
- [13] S. Uma. and J. Suguna, “Human Interaction Pattern Mining using Enhanced Artificial Bee Colony Algorithm”, *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 3, No. 9, pp. 2320-9798, 2015.
- [14] S. Uma and J. Suguna, “Temporal Semantic Analysis Based Human Interaction Pattern Mining using Partial Ancestral Graph”, *Research Journal of Applied Sciences, Engineering and Technology*, Vol. 8, No. 12, pp. 1487-1491, 2014.