Extracting frequent pattern from Human Interaction in Meeting using Tree based Approach

Prashant Puri and Kirti Korabu

Department of Information Technology, Sinhgad College of Engineering, Pune,India

Abstract

Human Interaction in meetings is one of the famous fields of social dynamics. Meeting is integral part of every organization. In this, meeting outcome is extracted using tree based approach. Meetings contents or conversation are available in forms such as audio, video and text. In this, pattern of meeting is extracted from text document. An interaction is represented in the form of tree. Meeting Output is generated using data mining technique. Firstly the contents are filtered, extracted and steamed. Secondly classification is done into six categories propose, comment, acknowledgement, request Info, ask Opinion, pos Opinion, and neg Opinion. Next the interaction tree is constructed which represent the interaction flow of meeting. Finally the meeting output is generated from interaction tree using frequent pattern mining algorithm. The behavior of person is determined which includes a person who proposed a lot of ideas, a person with positive or negative attitude.

Keywords

Social Dynamics, Stop words, Frequency based Classification, Frequent Pattern Mining, Human Interaction In Meetings

1.Introduction

Group social dynamics is one of the important areas in the field of research. Human interaction in meeting is one of the important characteristics of group dynamics. Group social dynamics is important for understanding the nature of meeting or understanding how conclusion was reached Meetings are important for purpose of information exchange, problem solving, knowledge sharing and creation. Meetings content a large amount of social and communication information.

Study of meetings explores the social behavior of participants in meetings which help us to understand the conclusion of meetings, whether all members agreed on outcome, who did not give opinion, who spoke little or lot. Extracted information from meetings is useful to predict the human interaction which is useful for meetings participants, meeting organizer, meeting sponsor.

In this paper, techniques for extracting information from meeting to predict the human interaction are discussed. Meeting contents are available in three forms audio, video and text which is used to predict the human interactions. To predict the human interaction feature extraction and classification of extracted feature is done. Feature extraction, classification, construction of interaction tree and determine frequent pattern for meeting are discussed in this paper on basis of text data.

The remainder of this paper is organized as follows. Section II discusses related work. Section III discussed the implementation of system. Section IV explores some applications. Finally, we conclude the paper in Section V.

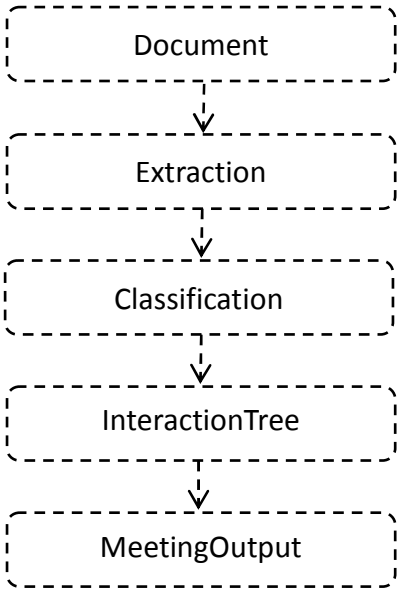
2.Related Work

Various researches have been conducted on discovering knowledge about human actions by applying the concept of data mining. Casas-Garriga [3] proposed algorithms to mine unbounded episodes from a sequence of events. The work is used to extract frequent episodes, i.e., collections of events occurring frequently together. Morita [5] proposed a pattern mining method for the interpretation of human interactions in a poster exhibition. It extracts simultaneously occurring patterns of primitive actions such as gaze and speech.I.McCowan[2] proposed the technique to detect the group level interest in meetings.Zhiwen Yu[6] adopted a multimodal method to infer human interaction based on a variety of features, such as speech tone, speaking time, interaction occasion, gestures, attention and information about the previous session. Four kinds of classification models, Support Vector Machine (SVM) uses LIBSVM, Bayesian Net, Naive Bayes, and Decision Trees [9] are selected to infer the type of each interaction. The results show that SVM is most appropriate and achieves a recognition rate of approximately 80 percent [1]. Please refer to the earlier paper [7] for details about the human interaction recognition. Zhiwen Yu proposed a tree base mining method that discovers patterns of human interaction in meetings [8]. Garg[4] proposed an approach to recognize participant roles in meetings.

3.Implementations of System

In this system, the meeting output is predicted form conversation of the participants. The conversations of participants are input to the system as text file. To predict the output flow of the meeting is identify and from the flow interaction tree is constructed. A frequent pattern mining algorithm is used to extract the output of the meeting. Also the meeting behavior and person behavior is determined from the system. To extract the content of meeting following are the measure steps

1. Extraction
2. Classification
3. Designing the Interaction tree.
4. Determining the meeting Output.



3.1EXTRACTION

In this step, contents of the meeting are filter to remove the stop words from statements in text file that contains the conversation. After removal of stop words, the steaming is performed to classify the words or phrase to the particular category.

a.Stop Word *:-* These are the words that some users leave out of an statement. By dropping stop words from an statement, the index size can typically be reduced by as much as 30% for a word level index. A stop word typically is a word which doesn't contain much "informational"content. For example, some example stop words are: "and", "the", "of", "it", "as", etc.

Eg : Statement before removing stop word
I think we should stage a concert to raise money.
After Extarction: * think * * stage * concert * raise money.

Here I, we, should,a and to is removed form statement.

b.Stemming :- A stemming is a process of linguistic normalization, in which the variant forms of a word are reduced to a common form, for example,

Connection
Connections
Connective ---> connect
Connected

3.2CLASSIFICATION

Classification Input to classification is filtered words from extraction. The extracted words from setp1 are classified into following categories propose, comment, Query, ask Opinion, pos Opinion, neg Opinion and acknowledgement. Classification is done using frequency based classification algorithm. Under each category there is list of word if word exits in a statement, the given statement is classified to a category to which that word belong.

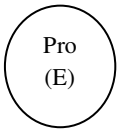
e.g1: Concerts take too long to organize.
As "too long" is negative phrase this statement will come under **negative category.**

e.g2: Yes tweets are very famous.
As "Yes" is positive word this statement will come under **positive category.**

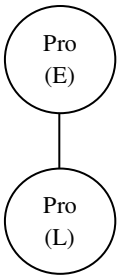
3.3 DESIGNING THE INTERACTION TREE

In this, flow of conversation is determined. The flow is identified once the interaction tree is constructed using the output of the classification step. This step will give the exact flow of interaction in the meeting and attitude of participants in the meeting. Tree will be constructed according to the following process.

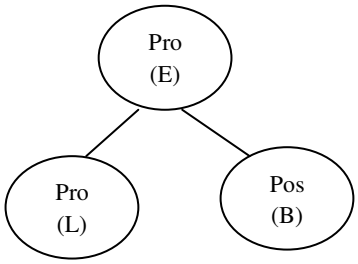
1st statement **Elena: I think we should stage a concert to raise money.** As it is a first statement, it is proposed statement. Here Pro (E) means Propose by Elena



2ndstatement **Lucas@Elena: Concerts take too long to organize. Let us have a bakesale.** Here Lucas@Elena means lucas reply to elena. As it is again a proposed statement and reply to first this statement will child of first. So the tree will be as follow



3rd statement **Barbara@Elena: My cousin is in a band called The Tweets that might play for free.** Here Barbara response to Elena**.** As it is a positive response to 1st statement, it will be other child to first.



3.4 MEETING OUTPUT-

The interaction tree is used to generate the meeting output. A frequent pattern mining algorithm is used to extract the frequent pattern from a tree generated from meeting interaction. A pattern is frequent trees or subtrees in the tree database. Following is the algorithm to find frequent tree and subtree.

Algorithm 1. fitm (TD, α) (Frequent interaction tree pattern mining)

Input: a tree database TD and a support threshold α
Output: all frequent tree patterns with respect to α
Procedure:
(1) scan database TD, generate its full set of isomorphic trees, ITD
(2) scan database ITD, count the number of occurrences for each tree t
(3) calculate the support of each tree
(4) select the trees whose supports are larger than α and detect isomorphic trees; if m trees are isomorphic, select one of them and discard the others
(5) output the frequent trees

Algorithm 2. fistm (TD; α) (Frequent interaction subtree pattern mining)

Input: a tree database TD and a support threshold α
Output: all frequent subtree patterns with respect to α
Procedure:
(1) i= 0
(2) scan database TD, calculate the support of each node
(3) select the nodes whose supports are larger than α to
form F^1
(4) i =i + 1
(5) for each tree t^i in F^i, do
(6) for each node t^1 in F^1, do
(7) join t^i and t^1 to generate C^{i+1}
(8) Subtree Support Calculating (TD; t^{i+1})
//calculate the support of each tree in C^{i+1}
(9) if there are any trees whose supports are larger than α, then select them to form Fiþ1 and return to Step (4)
(10) else output the frequent subtrees whose supports are Larger than α

Subprocedure. Subtree Support Calculating (TD, st)

(1) count=0
(2) supp(st)= 0
(3) for each tree t € TD do
(4) create subtrees S of t with any item s € S, |s| = |st|
(5) flag false
(6) for each item s €S do
(7) generate isomorphic trees IS of s
(8) for each item is € IS do
(9) if tsc(is)= tsc(st) then
(10) count =count + 1
(11) flag = true
(12) break
(13) if flag = true then
(14) break
(15) supp(st)= count/|TD|
(16) return supp(st)

Table 1 .Notation

Notation	Description
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- subtrees
A	A support threshold minsup

The support of T is defined as

$$supp(T) = \frac{number\ of\ occirences\ of\ T}{total\ number\ of\ trees\ in\ TD}$$

:

4.Applications

1. **Corporate Meetings**; - Meeting is integral part of every organization To analyze the meeting output is of great concerned for organization. So this system can help the company to predict the meeting output.
2. **Business Analyst**:- Business Analyst is always concerned for participant behavior in meeting. The participant behavior means who talks most time or who talks for very less time, who talks positively or negatively in meetings.
3. **Civil Court**:- This system is used for Lawyers to analyze the case study and accordingly study the different cases.

4. **Statistics Evaluation of Viewers for Reality Show**:- Viewers give their views for the show online for eg on facebook , tweeter so these views can be categories and result of viewers can be evaluate.

5.Conclusion

In this report, human interaction system is proposed which extract the contents of meeting and predicts the output of meeting. A tree based mining approach is used for discovering frequent patterns of human interaction in meetings discussion. For filtering the meeting contents stop words removal and steaming is applied. To classify the conversation simple classification technique is used. Frequent pattern mining algorithm is used to extract the frequent tree and meeting output is generated. Proposed system also predicts behavior of meeting and determined the behavior of participants. It determines the persons who proposed a lot of ideas, the persons who were critical, whether all members agreed on the outcome, who did not give his opinion, who spoke a little or a lot.

ACKNOWLEDGEMENTS

It is my pleasure to get this opportunity to thank my beloved and respected Guide Prof. K.S.Korabu who imparted his valuable knowledge specifically related to image processing. We are grateful to department of Information Technology SCOE, Pune for providing us infrastructure facilities and moral support.

REFERENCES

[1] G. Casas-Garriga, “Discovering Unbounded Episodes in Sequential Data,” Proc. European Conf. Principles and Practice of Knowledge Discovery in Databases (PKDD ’03), pp. 83-94, 2003.
[2] T. Morita, Y. Hirano, Y. Sumi, S. Kajita, and K. Mase, “A Pattern Mining Method for Interpretation of Interaction,” Proc. Int’l Conf.Multimodal Interfaces (ICMI ’05), pp. 267-273, 2005.
[3] D. Gatica-Perez, I. McCowan, D. Zhang, and S. Bengio, “Detecting Group Interest-Level in Meetings,” Proc. IEEE Int’l Conf. Acoustic, Speech, and Signal Processing, vol. 1, pp. 489-492, 2005.
[4] Zhiwen Yu; Zhiyong Yu; Aoyama, H.; Ozeki, M.; Nakamura, Y., "Capture, recognition, and visualization of human semantic interactions in meetings," Pervasive Computing and Communications (PerCom), 2010 IEEE International Conference on , vol., no., pp.107,115, March 29 2010-April 2 2010.
[5] Z.W. Yu, Z.Y. Yu, Y. Ko, X. Zhou, and Y. Nakamura, “Inferring Human Interactions in Meetings: A Multimodal Approach,” Proc. Sixth Int’l Conf. Ubiquitous Intelligence and Computing (UIC ’09),pp. 14-24, July 2009.
[6] Z.W. Yu, Z.Y. Yu, Y. Ko, X. Zhou, and Y. Nakamura, “Tree-based mining for discovering patterns of human interaction in meetings,” Proc.IEEE Transactions on knowledge and data engineering, vol. 24, no. 4, april 2012.
[7] Garg, N. P., Favre, S., Salamin, H., Tur, D. H., and Vinciarelli, A Role Recognition for Meeting Participants: an Approach Based on Lexical Information and Social Network Analysis. In Proc. ACM Multimedia 2008, 693-696.
[8] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin,”A Practical Guide to Support Vector Classification”
[9] Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines (2001), http://www.csie.ntu.edu.tw/~cjlin/libsvm

[10] Weka (2008), http://www.cs.waikato.ac.nz/ml/weka/ .
[11] Minqing Hu and Bing Liu."Mining and Summarizing Customer Reviews."Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004), Aug 22-25, 2004, Seattle,Washington, USA,

Author

Prashant P Puri He is student at Sinhgad College of Engineering, Pune. He received his bachelor degree in Information Technology in 2011, and currently pursuing M.E. His current research interest includes developing mining methods for inferring human interactions in meetings in the domain of Data Mining.

Kirti Korabu She received her bachelor degree and Master Degree in Computer science. She has a experience of 18.5 yrs and currently working as Associate Professor in Sinhgad College of Engineering. She has published five international journal papers. Her current research interest includes Data Mining, Information Retrieval, Software Engineering, Database Management Systems, Data Structures and Algorithms, Theory of computation. She is member of LMISTE, LMCSI.