IJSET International Journal of Scientific Engineering and Technology (ISSN : 2277-1581)
Volume No.2, Issue No.9, pp : 866-871 1 Sept. 2013

Tree-Based Mining with sentiment Analysis for Discovering Patterns of Human Interaction in Meetings

P. Maheswari, R. Manavalan
Department of computer science and Applications, K.S.Rangasamy college of Arts and Science
Thiruchengode, amilnadu, India
Email Id: manavalan_r@rediffmail.com, pmaheswari.22@gmail.com

Abstract-*Human interaction is one of the most important characteristics of group social dynamics in meetings. The sequence of human interaction is generally represented as a tree. Tree structure is used to capture how the person interacts in meetings and to discover the interactions. The human interaction are proposing as an idea, giving comments, ask opinion, acknowledge, etc., Frequent interaction tree pattern mining algorithm and Frequent interaction sub tree pattern mining algorithm are utilized to analysis the structure and to extract interaction flow patterns, where co-occurring only the tags are considered. To overcome this problem, Sentiment Analysis (SA) is proposed work to the entire flow of interaction in meetings. A sentiment analysis approach to extract sentiments associated with opinions of positive or negative for specific subjects the from the document instead of classifying the whole document into positive or negative. Sentiment analysis approach identifies the semantic relationship between the sentiment expressions and subject properly and also improve the performance of discovering pattern of Human interactions in meetings.*

Keywords - Tree based mining, frequent interaction sub-tree mining, frequent interaction mining and Modified Embedded sub tree mining.

I. INTRODUCTION

In the social context, human interaction is the most important factor for understanding the human behaviour or activities in the meeting and determines whether the meeting is well organized. Several methods have been proposed to find the flow of the interaction of each human being in the meeting. Further to understand the interference of human interactions in meetings, it is essential to discover higher level semantic knowledge such as frequent interactions flow that often occur during discussion and usually follows relationships among the exits interactions. This knowledge will help to describe important patterns of human interactions in meetings, it is essential to discover higher level semantic knowledge such as frequent interactions flow that often occur during discussion and usually follows relationships among the exits interactions. Frequent patterns of human interaction are extracted based on the captured content of face-to-face meetings. Human interaction flow in the meetings is defined as an idea of expressing a positive opinion, negative opinion and giving comments. The mining results can be used for indexing semantics of the meeting. Existing meeting capture systems could use this technique as a smarter indexing tool to search and access particular semantics of the meetings. Interaction tree pattern mining algorithms is used to analyze tree structures and extract interaction flow of patterns in the meeting.

Sentiment Analysis (SA) is also known as opinion mining. Opinion mining is a type of natural language processing for tracking the feeling of the public about a particular product. The sentiment found within comments, feedback or critiques provide useful indicators for many different purposes. These sentiments can be categorised either into two categories: either positive or negative; or into an n-point scale, e.g., very good, good, satisfactory, bad, very bad. A sentiment analysis task can be used as a classification task where each category represents a sentiment. The overview of the proposed model is shown in Fig 1.

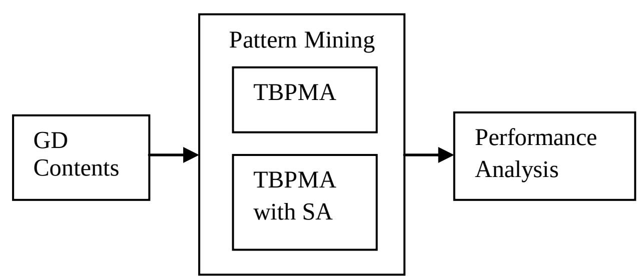

Fig 1.Overview of Proposed Model

The rest of this paper is organized as follows: In Section II, previous studies related to Tree Based Pattern Mining Algorithm (TBPMA) is discussed. Section III describes human interaction definition and recognition, as well as interaction flow instruction. The tree-based interaction pattern mining method is presented in Section IV. The Tree-Based interaction Pattern Mining method with Sentimental Analysis describes in Section V. The experimental result analysis and Discussion is presented in Section VI. Finally, the work is concluded in Section VII.

II. RELATED WORKS

Human interaction in meetings has attracted much research in the fields of image/speech processing, computer vision, and human-computer interaction [i]. Stiefelhagen et al. [ii] used microphones to detect the current speaker and combined acoustic cues with visual information for tracking the focus of attention in meeting situations. McCowan et al. [iii] recognized group actions in meetings by modeling the joint behavior of participants based on a two-layer Hidden Markov Model (HMM) framework. The AMI project [iv] was proposed for studying human interaction issues in meetings, such as turn-taking, gaze behavior, influence, and talkativeness. Otsuka et al. [v] used gaze, head gestures, and utterances in determining interactions regarding who responds to whom in multiparty face-to-face conversations. DiMicco et al. [vi] presented visualization systems for reviewing a group's interaction dynamics, e.g., speaking time, gaze behavior, turn-taking patterns, and overlapping speech in meetings. In general, the above-mentioned systems aim at detecting and visualizing human interactions in meetings, while work is focused on discovering higher level knowledge about human interaction.

Several works have been introduced so for in discovering human behavior patterns based on stochastic techniques. Bakeman and Gottman [vii] applied sequential analysis to observe and analyze human interactions. Magnusson [viii] proposed a pattern detection method, called T-pattern to discover hidden time patterns in human behavior. T-pattern has been adopted in several applications such as interaction analysis [ix] and sports research [x]. Although the purpose of these techniques is similar to the proposed work, an analysis on human interaction in meetings is conducted and the problem of discovering interaction patterns from the perspective of data mining was addressed. A number of studies have been conducted on discovering knowledge about human actions by applying the concept of data mining. Casas-Garriga proposed algorithms to mine unbounded episodes (those with unfixed window width or interval) from a sequence of events on a time line [xi].

The propose of TREEMINER algorithms also proposed to discover all frequent sub trees in a forest, using scope- list. TREEMINER with a pattern matching tree mining algorithm is called PATTERNMATCHER Tree-Based Pattern Mining algorithm are (A) only considered the small dataset. (B) It is not used for the quality of medical interview leading to patient's satisfaction, effective medical treatment and so on. (C) Does not consider the verbal behaviors. (D) It includes only the simple nonverbal behaviors. (E)This is applicable for only low-level features. To address issues of nonverbal behaviors, Sentimental Analysis Algorithm is integrated with Tree Based Pattern mining Algorithm for classifying the document.

III. HUMAN INTERACTION AND ITS FLOW

The meaning of human interaction varies depending on the usage of the meetings or the types of the meetings. This research focuses mainly on the task-oriented interactions. The other communicative actions that concern the meeting and the group [iii] itself (e.g., when someone invited another participant to take the floor) are not included. A set of interaction types based on a standard speech-unit tagging scheme is created such as propose, comment, acknowledge, request information, ask opinion, positive opinion, and negative opinion. Person's proposes an idea with respect to a topic. For Example one job selection meeting (10 min, talking about factors that will be considered in seeking a job, such as salary, working place, employer, position, interest, etc.).

3.1. 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 among them. Fig 2 shows Tree Representation for Human Interaction Flow.

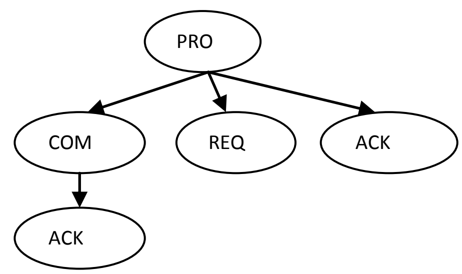

Fig 2.Example Tree Representation for Human Interaction Flow

For this representation used labels for human interactions. L = {PRO, COM, ACK, REQ, ASK, POS, NEG. Then, the Tree-Based interaction pattern mining is described in Section 4.

IV. TREE-BASED INTERACTION PATTERN MINING

Mining frequent patterns is probably one of the most important concepts in data mining. A lot of other data mining tasks and theories stem from this concept.

4.1Tree-Based Pattern Mining Algorithm (TBPMA)

Tree-Based Pattern Mining algorithm is designed for interaction flow. Frequent tree pattern mining algorithm [x] is used for each tree in TD, for exchanging the places of siblings (i.e., performs commutation processing) to generate the full set of isomorphic trees, ITD. The tree dataset is constructed by using String Encoding method which is explained under here.

String Encoding

String Encoding defines an Application Programming Interface (API) for encoding strings to binary data, and decoding strings from binary data. To represent a Tree Dataset (TD), String Encoding method is used. The first node of the root is represented using "-", and sub tree is represented by using "()", sibling relationships are represented using "*". The example string code for Fig1 tree is PRO-(COM-ACK)*REQ*ACK. First, tree dataset (TD) is constructed. Then Tree based mining algorithm calculates the support of each node (tree) in ITD. Then it selects the trees whose supports are larger than σ (minsupport). It finally outputs the frequent patterns as frequent trees. Frequent interaction tree pattern mining Algorithm (FITPM) is projected in Fig 3.

Frequent Interaction Tree Pattern Mining Algorithm
Input: a tree database TD and a minsupport σ
Output: all frequent tree patterns.

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.

Where, TD - A dataset of interaction trees.
ITD - The full set of isomorphic trees to TD
T - A Tree ,σ - A support threshold minsup

Fig 3. Frequent Interaction Tree Pattern Mining Algorithm

For each tree in TD, to generate full set isomorphic trees. The purpose of generating isomorphic trees is to ease string matching. It then calculates the support of each tree in ITD; it selects the trees whose supports are larger than σ and detects isomorphic trees with in them. If m trees are isomorphic, it selects one of them and discards the others. It finally outputs the frequent trees. Frequent Interaction Sub Tree Pattern Mining Algorithm (FISTPM) is projected in Fig 4.

Calculate the support of each node and selects the nodes whose supports are larger than σ to form the set of frequent nodes. Then adds a frequent node to existing frequent i-sub trees to generate the set of candidates with i + 1 nodes. If there are any trees whose supports are larger than σ, it selects them to form F^{i+1} and repeats the procedure. Otherwise, it stops to output the frequent sub trees.

Given a tree or sub tree T and a data set of trees TD, the support of T is defined as in equation 1.

$$\text{supp}(T) = \frac{\text{Number of Occurrences in T}}{\text{Total Number of Trees in TD}} \quad (1)$$

If the value of supp (T) is more than a threshold value minsupp T, It is called a "frequent tree" or" frequent sub tree."

Frequent Interaction Sub Tree Pattern Mining Algorithm
Input: a tree database TD and a support threshold σ
Output: all frequent sub tree patterns with respect to σ

Step1: Initialize the value i = 0
Step2: Scan database TD, calculate the support of each node
Step3: Select the nodes whose supports are larger than σ to form F^1
Step4: Then add a frequent node to existing frequent i-sub trees to generate the set of candidates with i+1nodes.
i= i+1
for each tree t^i in F^i ,do
for each node t^1 in F^1,do
join t^i and t^1 to generate C^{i+1}
Step5:calculate the support of each tree in C^{i+1} (Subtree_Support_Calcultating(TD, t^{i+1})
Step6: If there are any trees whose supports are larger than σ, then select them to form F^{i+1} and return to Step (4).
Step7 :else output the frequent sub trees whose supports are larger than σ

Where, t^k - A sub tree with k nodes, i.e., k-sub tree
C^k - A set of candidates with k-nodes,
F^k - A set of frequent k-sub trees

Fig 4.Frequent Interaction Sub Tree Pattern Mining Algorithm

V. Tree Based Interaction Pattern Mining With Sentiment Algorithm

Sentiment Analysis (SA) is a type of subjectivity analysis which aims to identify opinions, emotions and evaluations expressed in natural language processing. The main goal is to predict the sentiment orientation (i.e. positive, negative or neutral) by analysing sentiment [xiii] or opinion words and expressions in sentences and documents. It has been studied by many researchers in recent years. Sentiment analysis approaches often require resources such as sentiment lexicons [xii] to determine which words or phrases are positive or negative in a general context.

Generally, there are three types of sentences in the English language: simple sentence, compound sentence and complex sentence. Simple sentences [xv] contain only one independent clause whereas compound sentences contain two or more independent clauses, and complex sentences contain at least one independent clause and one dependent clause. In structure, a clause comprises a subject and a predicate, where the predicate is a combination of verb, object, complement and adverbial. The subject is usually a noun phrase that names a person, place or thing. The verb identifies an action or a state of being. An object receives the action and usually follows the verb. SA is an application of Natural Language Processing (NLP).It deals with the actual text element. It transforms it onto a format that the machine can use. Other use of SA is Artificial Intelligence (AI). It uses the information given by the NLP and uses a lot of methods to determine whether something is

International Journal of Scientific Engineering and Technology (ISSN : 2277-1581)
Volume No.2, Issue No.9, pp : 866-871 1 Sept. 2013

positive or negative, for clustering. The clause level sentimental analysis algorithm is presented in Fig 5.

Clause Level Sentiment Analysis Algorithm **Input:** our own dataset **Output:** sentiment expressions of given subject
Step 1: For each sentence of the review text, semantic annotation is performed, and prior sentiment scores are assigned to each word. Step 2 :The grammatical dependencies are determined, and the sentence is broken into independent clauses. Step 3: The contextual sentiment score is calculated by traversing the dependency tree based on its clause structure. Step 4: Processing all the clauses of the sentence. Step 5: The sentiment score for each review aspect is calculated by taking the average of the clause addressing the same aspect. Step 6: The sentiment score for the whole sentence is determined from the calculated scores of multiple Aspects.

Fig 5. Clause Level Sentiment Analysis Algorithm

The task of sentiment analysis is to find sentiment expressions for a given subject and determine the polarity [xiv] of the sentiments. In other words, it is to identify text fragments that denote a sentiment about a subject within documents rather than classifying each document as positive or negative towards the subject.

5.1 Framework of Sentiment Analysis

Besides adjectives, other content words such as nouns, adverbs, and verbs are also used to express sentiments. In principle, a sentiment expression[xii] using an adjective, say "*good*", denotes the sentiment towards its modified noun such as in "*good product*," and the whole noun phrase ("*good product*") itself becomes a sentiment expression with the same polarity as the sentiment adjective (positive for "*good*" in this case). Likewise, a sentiment expression using an adverb, say "*beautifully*," denotes the sentiment towards its modified verb such as in "*play beautifully*," and the polarity of the sentiment is inherited by the modified verb. Thus, sentiment expressions using adjectives, adverbs, and nouns can be simply defined as either positive or negative in terms of polarity.

- Sentiment verbs that direct either positive or negative sentiment toward their arguments,
- Sentiment transfer verbs that transmit sentiments among their arguments, and associate them with arguments such as subjects and objects that inherit or provide sentiment.

Notations	Descriptions
Polarity	Positive (good), negative (bad), or neutral is denoted by **g**, **b**, or **n**, respectively, and sentiment transfer verbs are denoted by **t**.
Part of speech(POS)	Currently, adjective (**JJ**), adverb (**RB**), noun (**NN**), and verb (**VB**) are registered in our lexicon.
Sentiment term in canonical form	Arguments such as subject (**sub**) and object (**obj**) that receive sentiment from a sentiment verb.

Fig 6.A simple notation of sentiment expressions

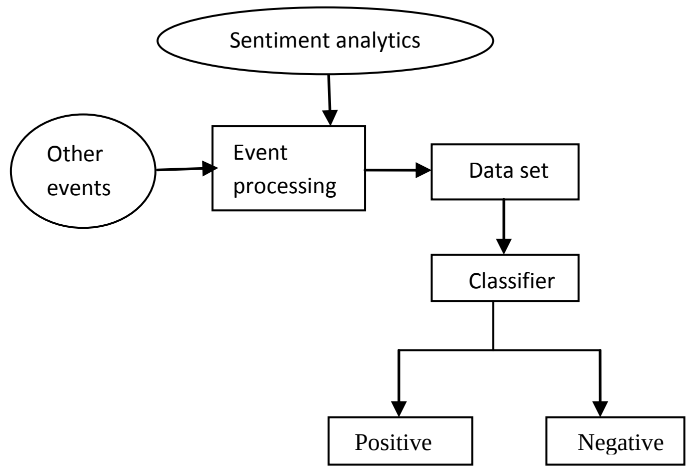

Fig 7. Sentiment analysis framework

Sentiment Analysis frame work is projected in Fig 7. The inputs are given by the GD contents (dataset). The classifier (sentiment lexicon) [xii] is used to classify the sentence into positive or negative. Finally the patterns such as Acknowledgement, Comments, and Neutral results are evaluated.

VI. EXPERIMENTAL ANALYSIS& DISCUSSION

The goal of experimental analysis is to calculate the accuracy of the proposed SA algorithm for interaction in meetings. Parameter that can be used for evaluating in this proposed model is sentiment expressions from the participants. The detailed description of dataset used for experiment and analysis of experiment; results and their discussions are mentioned below.

6.1Dataset

This work involves real meetings lasting 15 minutes on average. Video camera, microphones and motion sensors used were for capturing the interaction meetings. Each meeting had four participants seated around a table. In order to use a correct data for mining, the interaction types are tuned manually after applying the recognition method. The goal of this work is to discover frequent interaction trees and analyzing the behavior of the algorithms on the data set, focusing on the effect of threshold. Fig8 and Fig 9 represent the tree and interaction tree.

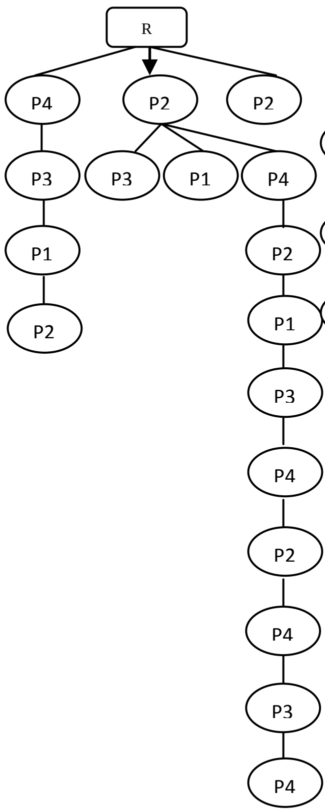

Fig 8. Representation of Tree

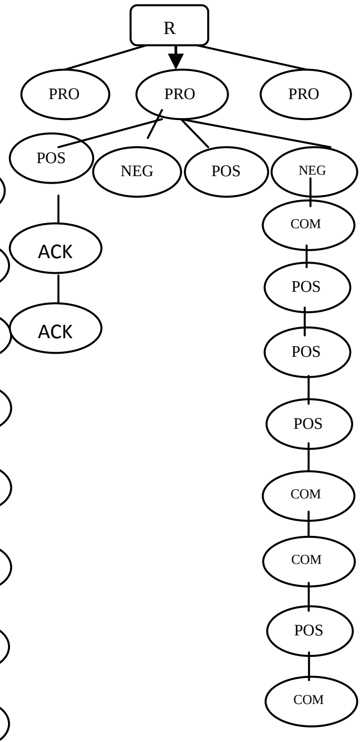

Fig 9. Representation of Interaction Tree

6.2Experimental Results and Discussions

The text files are used in order to evaluate the performance of proposed algorithm. The accuracy is a parameter; it is used to evaluate the efficiency of proposed sentiment analysis algorithm over the tree based pattern mining algorithm. The accuracy is defined on the sentiment expressions from the participants. Accuracy is defined in equation 2.

$$\text{Accuracy} = \frac{\text{no. of occurrences}}{\text{Total no. Of occurrences}} \quad (2)$$

From the Fig 8 p_1, p_2, p_3 and p_4 represents person1, person2, person3 and person4. In the Fig 9, R represents the root, PRO represents propose, POS represents positive opinion, ACK represents Acknowledgement, NEG represents Negative Opinion, COM represents comment. In the above Interaction Tree Hierarchy positive opinions are verbalized 5 times, Proposing is verbalized 3 times, acknowledgements are verbalized 2 times, comments are verbalized 2 times and NEG negative opinions are verbalized 2 times.

The text files have been taken from the internet and given input of six text files the text data has to go through POS' tagging pre-processing in order to classifies the vocabularies as verb, subject, object and so on. The test text files are considered for evaluating the proposed system that are simple enough to improve the proposed algorithm.

The text files have been taken from the internet and given input of six text files the text data has to go through POS' tagging pre-processing in order to classifies the vocabularies as verb, subject, object and so on. The test text files are considered for evaluating the proposed system that are simple enough to improve the proposed algorithm.

Table1 shows the computational results of Sentiment Analysis over the Tree Based Pattern Mining Algorithm. From the table 1 the accuracy of Tree Based Pattern Mining Algorithm for positive content is 8.9% where as 9.6%maximum accuracy is achieved by Sentiment Analysis Algorithm and it is 0.7%more than the Tree Based Pattern Mining Algorithm.

Table 1.Test Data and Average

Doc name	Positive content		Negative content		Comments		Acknowledg -ement	
	TB PMA	SA	TB PMA	SA	TB PMA	SA	TB PMA	SA
Ext. txt	12.76	14.5	13	14	59.8	63	9.2	9
Inp1. txt	8.7	8.5	21.4	20	60.5	62	9.5	9
Inp2. txt	12.3	13	18.0	19	57.5	58	10.7	10
Inp3. txt	12.5	13	19.5	19	57.2	58	9	10
Inp4. txt	1.3	3.4	17.1	18.3	76.5	76	2	2.8
Inp5. txt	6.4	5.6	18.5	19.7	69	70	3.1	3.2
Avg	8.9	9.6	17.5	18.3	63.4	64.5	7.1	7.3

Likewise the accuracy of Negative content in Tree Based Pattern Mining Algorithm is 17.5% where as the maximum accuracy of SA algorithm is 18.3%. It is 0.8% higher than Tree Based Pattern Mining Algorithm. The accuracy of Comments in tree based pattern mining algorithm is 63.4% when it compare to SA algorithm, which achieves 64.5% of accuracy. SA algorithm achieves 1.1% accuracy higher than Tree Based Pattern Mining Algorithm. The maximum accuracy of an Acknowledgement in Tree Based Pattern Mining Algorithm is 7.1% where as 7.3% of maximum accuracy is produced by Sentiment Analysis algorithm. SA algorithm achieves accuracy of 0.2%more than the Tree Based Pattern Mining Algorithm.

The experiment results of both Tree based pattern mining algorithm and SA algorithm calculated (i.e. positive content, Negative content, comments and Acknowledge) in table1. The computational results clearly explain that the TBPMA with SA algorithm is batter then Tree based pattern mining algorithm.

Fig 10. Performance Analysis of Tree Based Interaction Pattern Mining With Sentiment Algorithm

VII. CONCLUSIONS

This paper proposes an Sentiment Analysis with Tree Based Pattern Mining method for discovering frequent interaction. With this approach, to automatically identify the contextual polarity for sentiment expressions, achieving results that are significantly better than Tree Based Pattern Mining algorithm. The experiment is conducted over six text files. From the experimental results, the accuracy of SA with TBPMA is better than compared to TBPMA. In the future will adopt for several applications based on the discovered patterns.

REFERENCES

i. *Z. Yu and 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.*

ii. *R. Stiefelhagen, J. Yang, and A. Waibel, "Modeling Focus of Attention for Meeting Indexing Based on Multiple Cues,"IEEE Trans. Neural Networks,vol. 13, no. 4, pp. 928- 938, July 2002.*

iii. *I. Mccowan, D. Gatica-Perez, S. Bengio, G. Lathoud, M. Barnard, and 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.*

iv. *A. Nijholt, R.J. Rienks, J. Zwiers, and D. Reidsma, "Online and Off-Line Visualization of Meeting Information and Meeting Support," The Visual Computer: Int'l J. Computer Graphics, vol. 22, no. 12, pp. 965-976, 2006.*

v. *K. Otsuka, H. Sawada, and J. Yamato, "Automatic Inference of Cross-Modal Nonverbal Interactions in Multiparty Conversations," Proc. Int'l Conf. Multimodal Interfaces (ICMI '07),pp. 255-262, 2007.*

vi. *J.M. DiMicco, K.J. Hollenbach, A. Pandolfo, and W. Bender, "The Impact of Increased Awareness while Face-to-Face," Human-Computer Interaction, vol. 22, no. 1, pp. 47-96, 2007.*

vii. *R. Bakeman and J.M. Gottman, Observing Interaction: An Introduction to Sequential Analysis. Cambridge Univ. Press, 1997.*

viii. *M.S. Magnusson, "Discovering Hidden Time Patterns in Behaviour-Patterns and Their Detection," Behaviour Research Methods, Instruments and Computers, vol. 32, no. 1, pp. 93-110, 2000.*

ix. *L. Anolli, S. Duncan Jr, M.S. Magnusson, and 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.*

x. *K. Gudberg, M. Jonsson, T. Anguera, P. Sa´nchez-Algarra, C. Olivera, J. Ssantos, J. Campanico, M.Castaner, C. Torrents, M.Dinusova, J. Chaverri, O. Camerino, and M.S. Magnusson, "Application of T-Pattern Detection and Analysis in Sport Research, "Open Sports Sciences J., vol. 3, pp. 95-104, and 2009.*

xi. *Ramakrishna, Rakesh Agrawa (1996)"Generalization and performance improvements".*

xii. *M. Collins. 1997. Three generative, lexicalized models for statistical parsing. In ACL-1997.*

xiii. *P. Beineke, T. Hastie, and S. Vaithyanathan. 2004. The sentimental factor: Improving review classification via human-provided information. In ACL-2004.*

xiv. *K. Dave, S. Lawrence, and D. M. Pennock. 2003. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In WWW-2003.*

xv. *V. Hatzivassiloglou and K. McKeown. 1997. Predicting the semantic orientation of adjectives. In ACL-1997.*