Analysis of coordinating activities in Collaborative Working Environments

Muhammad Muneeb Kiani, Abdul Rafay Awan

University of Milan, Milan Italy, Teradata, Islamabad Pakistan
Muhammad.kiani@unimi.it, abdul.rafay@teradata.com

Abstract. Collaborative Working Environments (CWE) are widely used for effective collaboration among users. A CWE includes various tools and methodologies to support analysis of coordinating activities. Users within a CWE widely utilize textual means of collaboration and communication. An effective analysis of this textual collaboration can help in improving overall quality of collaborating activities and monitoring of the CWE. In textual analysis text is analyzed within certain context. Existing semi-automated techniques which are based on lexical, syntactic, semantic and other analysis approaches can be utilized with addition of customized automated classifier to cater needs of analyzing coordinating activities in collaborative working within a specific context. In proposed framework, natural language processing, opinion mining, lexicon based approaches will serve as processors of the framework.

Keywords: Text Analysis, Collaborative Working Environments

1 Introduction

Advancements in technology made possible interactive communication with computer systems and among other users. Further down the road with the help of these new technologies paradigm of collaborative working emerged. Set of tools that supports notion of collaborative working are known as Collaborative Working Environment (CWE) [**1**]. Part of coordinating activities in collaborative working environment includes textual communication. Though there are various techniques available for monitoring collaborative working but monitoring textual communication is still a complex task.

NLP (Natural Language Processing) can offer relevant support for the automatic classification of actions. In recent years researchers developed various mechanisms and algorithms for analyzing text. The term text analysis refers to the tasks that are performed to interpolate the facts and figures to augment the decision making and predicting future trends. There are two categories of text analysis; first category is the analysis of structured data that is performed on the data warehouse of an organization to find out different statistics of a business whereas the second category is the analysis

of unstructured data i.e. web logs, audios, videos, etc. to predict the market trend and what are the reasons for the failure of a particular product etc.

Text analysis techniques can be useful in monitoring coordinative activities within a CWE, understanding coordination among users in a given CWE and context can greatly enhances overall effectiveness of the systems. Though textual analysis techniques mentioned earlier are very mature, but they require further customization in context of CWEs for an accurate analysis.

In this paper a model is presented to analyze the collaboration within a certain context by using existing text analysis techniques to augment the CWE.

2 Background

The increasing degree of connectivity has given rise to the collaborative environments. Governments and corporations have adapted to networked collaborative environments to deliver their services. Managing the ever changing dynamics of collaborative environments and putting in place effective monitoring processes has become an important competency parameter. In order to determine the quality index of text based collaborative environments use of NLP is the inherent choice. Lexicon based techniques were used to analyze the activity model as proposed by the activity theory to analyze and identify the cognitive advantages of joint activity [**2**].

As mentioned earlier in section 1, CWE utilizes textual communication means for coordinating activities. These activities are performed with the help of rich text editors, group chat messages, emails and other means. This involves a lot of textual data, and effective understanding and monitoring of this data can greatly help in improving systems and overall activities from various aspects. Textual analysis is one way to understand this information. Following is small brief of various textual analysis methods.

Some research work has been done in other languages for word sense disambiguation [**3**]. Basing on the work of WordNet, Esuli and Sebastiani introduced another library named as SentiWordNet for the purpose of opinion mining. SentiWordNet utilize lexical basis of WordNet and assign certain value to different words in terms of positivity, negativity and neutrality which as result help in determining overall mood of a textual data [**4**].

3 Need for Analysis of coordinating actives in CWE

Lyk et al explained role of monitoring in a CWE in various stages [**5**]. Great deal of work has been done and various open source libraries of text mining, classifying and analyzing text have been made available in recent years. Opinion mining or sentiment analysis determines overall mood of the textual data focusing on particular set of activities within a data set [**6**]. Isabella et al. analyzed coordinating activities against a predefined set of parameters [**7**]. But classification of individual items has been done

manually without a particular rule set so it is not possible to scale up or use this approach in similar scenarios. Secondly slicing of text has been done based on time stamp and later rectified manually with human intervention which is another bottle neck in this approach.
There is a need to better understand set of activities within context of a collaborative working environment which involves overall goal of collaboration such as brainstorming, surveys, coordinating activities, number of participant, timing etc. In order to analyze coordinating activities in a CWE exiting text analysis techniques can be utilize.

4 Proposed methodology

In proposed methodology coordinating activities of users will be recorded. For this purpose Innovation Factory [8] CWE will be used as it provides event logs of the coordination. Various textual analysis techniques will be tested and one that fits best for the purpose will be adopted. [7] Manually labeled a set of coordinating activities as shown in the Table 1. This data along with manually labeled data from further experiments will be used to evaluate the output of analysis.

Type	Originator	Text Token	Unit
Group-Chat	Samo Rumez	where are you all?	Situation Request
Group-Chat	Samo Rumez	no, i think Vesna is eating	Situation Update
Group-Chat	Vesna Paulic	ok guys lets start	Plan Propose
Text	Nikolaj Potocnik	we can open all tv channels for one week before	Lay outing
Poll Question	Primoz Klasinc	Campaign should be about	Information/knowledge Request
Poll Vote	Nikolaj PotoÄ • nik	Standalone Mobia	Information/knowledge Provision

Table 1. Short Snapshot of Manually labelled Data

Architecture of proposed methodology comprises of three main sections, input, output and processor as depicted Fig.1 Input section provides Information about Classes; currently we have assumed four collaborating classes that are Query (sub class counter query), Opinion, Agreement, Argument (sub class counter argument).

More classes can be elicited depending upon the context such as classes defined by [**7**] includes (idea) generation, agreement, disagreement, neutral and coordination. Second input includes raw data set which should also include other value added information for comprehensive analysis such as timestamps, number of users, their input text, overall topics of discussions. Third and last input is set of quality metrics which helps analysis to assign weight to various activities based on type of context for example quality metrics will be different in case of a brainstorming, group discussion session than that of a question answer session or survey.

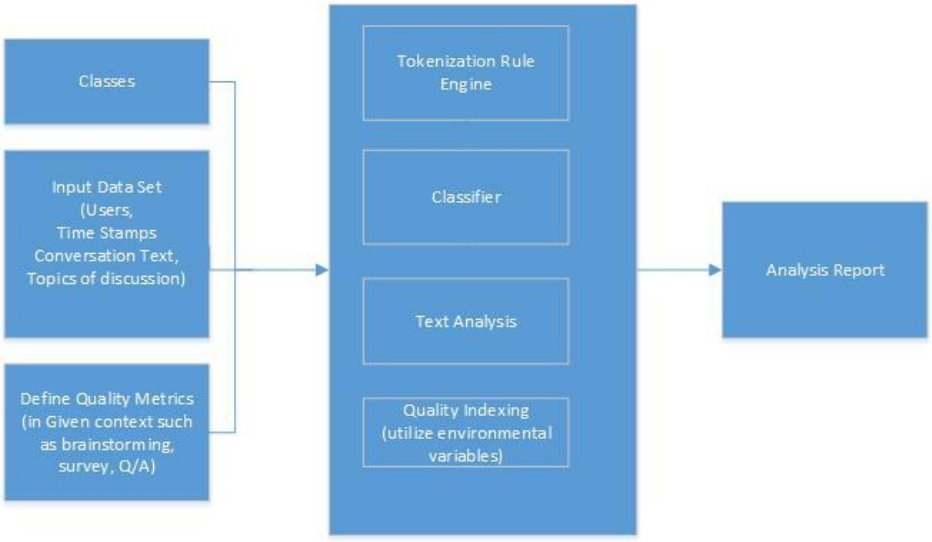

Figure 1: Overall Architecture of Proposed Methodology

Second section includes tokenization rule engine which determine how text should be sliced for analysis purposes, it necessary to correctly tokenize set of text for correct semantic linking. Tokenizing for finding queries is easy but finding other elements is complex set. Determining a generic rule engine requires extensive evaluation in various settings.

As discussed earlier rule of classifier is to link tokenized text for creating meaningful information, in case of coordinating activities larger set of text are required to be classified to correctly determine their relevant classes. There are various libraries available which provides support text analysis such as dandelion which can be used for analysis [**8**]. Dandelion analyzes text with respect to context and also provides API for further customization and allows extraction of various kind of information. Last component of processer comprises of various environment variable such as time stamps, user participation, topic of discussion and related these variables with textual information in order to determining overall quality of entire activity.

Output section is set of reports which are produced after processing raw data and contextual information.

5 Conclusion and Future Work

Proposed methodology provides basis for developing a comprehensive framework and tool support for CWE. For the labeling of data, crowdsourcing tool will be developed. Results will be evaluated to mitigate the under-fitting or over-fitting of classifiers that can create false positives. Once the classifier is trained and tested against the raw data, in the third step it will be incorporated in a CWE such as Innovation Factory [**8**] analysis and monitoring framework.

References

1. Georgia, B., Gregoris, M.: Review and functional classification of collaborative systems. International Journal of Information Management 22(4), 281–305 (2002)
2. Barros, B., Verdejo, M.: Analysing student interaction processes in order to improve collaboration. The DEGREE approach. International Journal of Artificial Intelligence in Education 11(3), 221--241 (2000)
3. Jianyong, D., Yao, F., Xia, l.: Attribute knowledge mining for Chinese word sense disambiguation. In : International Conference on Asian Language Processing (IALP), Suzhou, pp.33-77 (2015)
4. Andrea, E., Fabrizio, S.: SENTIWORDNET: A Publicly Available Lexical Resource. In : Language Resources and Evaluation, LREC 2006, GENOA (2006)
5. Harshada, P., Michael, P., John, R. .: Factors of collaborative working: A framework for a collaboration model. Applied Ergonomics 43(1), 1–26 (2012)
6. Pang, B., Lillian, L.: Opinion mining and sentiment analysis. Foundations and trends in information retrieval 2(1-2), 1-135 (2008)
7. Isabella, S., Ronald, M., Paolo, C., Fulvio, F.: Tracing the development of ideas in distributed, IT-Supported teams during synchronous collaboration. In : ECIS 2014 (2014)
8. Bellandi, V., Ceravolo, P., Damiani, E., Frati, F., Maggesi, J., Zhu, L.: Exploiting Participatory Design in Open Innovation Factories. In : Eighth International Conference on Signal Image Technology and Internet Based Systems (SITIS), pp.937--943