A REVIEW ON AUTOMATIC RESPONSE PROVISION FOR IMPROVING CONVERSATION

[1]Pranav Pathak, [2]Vijay Gujar
[1,2]Arvind Gavali College of Engineering, Satara,
[1]pranavpathak.rgpm@gmail.com, [2]gujar.vijay@gmail.com

Abstract- The addition of matter in the response to the answer while having conversation improves the engagement of people who are having conversation. This technique of adding additional information can consist of adding responses automatically. The responses can be expressive in order to make conversation enjoyable. Normally additional information is to extend the conversation by adding some data at the tail of answer. But that sometimes does not tempt to involve other person. But with the technique of automatic response provision, the expressive opinion will be added to the response. The responses are developed from large no. of reviews on the internet. Generating opinions also includes ranking the web results and selecting top ranked text. Also it adds phrases to the content if required in order to improve communication.

Keywords: Grice Maxim, Natural Language Processing, Sentence Generation, Small Talk.

I. INTRODUCTION

Consider an example where two persons are chatting

A: Today is sunny day.

B: Yes. It was very overcast yesterday. But today atmosphere looks warm, is not it?

Here person A simply states the current scenario. Then person B responses back with some additional information. Also person B asks the question along with information. This helps extending the conversation. Such kind of procedure is helpful in developing conversation between two distant people. If Grice Maxim [1] is considered, then the responses should not contain any extra bit of information. So in above conversation, according to Grice's Maxim too much information is present in the response generated by person B. But in day to day conversations more information and questioning approach are used for developing interests in the conversation process. The phenomenon called Small talk[2] is followed
for including the additional opinions in order to provide the content for continuing the conversation. Schneider [3] started working on small talk. His study concluded that taking initiative on topic discussion, adding phrases, responding to the content are the important factors in small talk conversations.

The generality level of expression decides the information within each sentence. To represent this level adjectives' frequency in documents and the number of adjectives in each sentence are considered. Also for this representation relevance of nouns to the current context, and each sentence's lengths are required. To generate opinions most of the times Natural Language Processing methods are considered. These methods include the processes such as sentiment analysis [4][5] and opinion mining [6]. Automatic opinion generation technique is adds an additional phrasing skill and generates automatic expressive opinion. The system typically behaves as follows: When a system receives any question from opposite, it first responses with appropriate answer. Then it adds another expressive opinion with the answer. This increases the possibility of engaging people in conversation more likely.

II. RELATED WORK

Opinion creation and assessment of sentiments is related to NLP which aims at recognizing subjective information from source material. This technique is basically used to determine the approach and thinking of user. It helps in deciding whether user's attitude is positive or negative. A sentiment analysis approach derives

the positive, negative and neutral polarities of input information. Turney[4] presented an approach of unsupervised learning for categorize the reviews. Pang[5] worked on the classification of documents on the basis of polarity. Yoichi Matsuyama et al.[11] has mentioned factoid based opinion generation scheme using natural language processing. Kamps *et al.* [8], [9] proposed a distance measure for the semantic orientation of adjectives. There is some contextual data around the subject term. The use of context information was proposed by Nasukawa *et al.*[10]

III. OPINION MAKING

Opinion making procedure consists of mainly four parts: Collection of documents, extraction of opinion, sentence style conversion, sentence ranking. In the document collection, phenomenon of small Talk is followed. In this, particular domain is selected. Then the reviews are observed and arranged as per ratings.

After that top 1000 reviews are considered In opinion extraction there are two processes: 1) Extraction of expressive responses and 2) Sorting of their opinion polarities. Here the negative reviews are eliminated. There are some particular words like right , wrong, good, bad etc. they clearly indicates the positive or negative sentiment. But there are some words like comfortable, affordable. For such words subjective evaluative dictionary is referred.

In sentence style conversion, the style of sentence is reworked. It simply includes an expression or question at the end. In sentence ranking the ranking scale is used to select the top reviews. Here the ranking scale contains the three components. (1)The magnitude of the word at which sentence correlates the topic.(2)Adjective frequency is the scale for unexpectedness. (3) The scale of morphemes. There are three rankings for algorithm : short , standard, diverse.

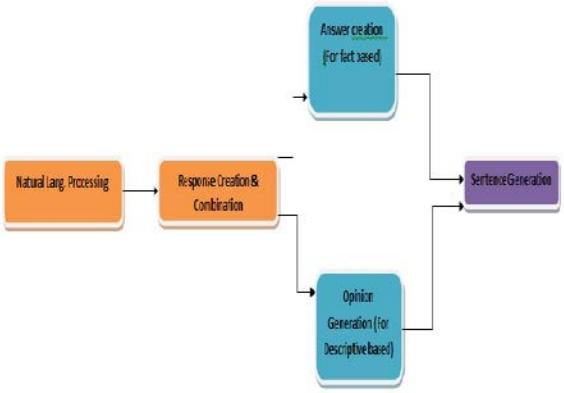

Fig.1 System Design

IV.SENTENCE RANKING

For expressing clear opinions the sentences should contain around seven to 10 morphemes. A **Short** algorithm generates short sentence as an opinion.

After filtrations based on topic coherence and no. of morphemes, sentences got sorted in terms of TF-IDF scores. Here top 30% sentences are considered.

TF-IDF consists of term frequency and inverse document frequency. It is calculated as follows:

$$TF(w,d) = \frac{C(w,d)}{\sum_{w} C(w,d)} \quad (1)$$

$$IDF(w) = \frac{|D|}{|\{d \in D : w \in d\}|} \quad (2)$$

$$TF - IDF(w) = TF(w,d) \cdot IDF(w) \quad (3)$$

where $C(w,d)$ represents the frequency of term w in a document d, $|D|$ represents the total number of documents in the corpus, $|\{d \in D : w \in d\}|$ represents the number of documents in which the term w appears at least once $(TF(w,d) \neq 0)$.

Fig. 2 TF-IDF

In next phase, the top 30% sentence list is arranged according to adjective frequency. For short sentences having more than two adjectives causes redundancy to give opinion. Here At this point, sentences consist of only one adjective are taken into consideration. For standard algorithm, top 30% list of sentences that are sorted on TF-IDF scores is extracted. These sentences consist of fifteen to twenty morphemes with single objective. If the sentence is containing the more than two nouns, the biggest TF-IDF is employed and adjective frequency sorts the 30% list.

Diverse algorithm extracts list of top 30% sentences that are having fifteen to twenty morphemes with multiple adjectives. In next phase this list is sorted according to the inverse order of averaged adjective frequency.

V. SYSTEM ARCHITECTURE

The Natural Language Processing is the initial point of the described system. Next is the response creation and combination process. The end part consists of fact based answers along with opinion sentence.

The input text analysis and conversion to question based text are the two parts involved in the natural language processing. The response creation and combination phase is again having two parts. Here in this phase it is decided that whether question is having fact based answer or descriptive answer. The fact based answers are specific and generated using semantic web techniques. Finally with combination of answer to particular fact oriented question and the opinion derived from the reviews by analysis based procedure the final response is generated.

A. Natural Language Processing:

In Natural Language processing, there is an interpretation of word. Each word is treated in terms of question mark and predicate. A hand crafted dictionary is used for such interpretation. Then next is topic estimation process. The topic estimation process is divided into 3 parts: morphological analysis of language, filtering of important words, and classification. Language morphological analysis separated nous from input text. Then in the next step important nous are collected from each topic based on the score of TF-IDF for each topic. Here top 50 important words are considered. Then linear chain conditional random fields (CRF) technique is followed for classification.

B. Response Creation and Combining Process:

For the creation of sentence, the process consists of two phases. First one is factoid type answer generation and the other one is opinion generation. The factoid type answers are specific and generated from structured databases.

C. Fact based Sentence Generation:

Semantic web techniques are used to generate factoid type responses. For this DBpedia project is referred. In this procedure structured contents are extracted from the data provided by Wkikipedia. The use of RDF eases representation of the extracted data. For extracting the factual information from Wikipedia pages, DBpedia is used. The RDF helps the users by creating statements in expressions according to subject, predicate and object. The subject is considered as resource. The predicate is aspect of resourse. The object indicates the entity that is in correlation with subject.

Next is question analysis procedure. Here it is determined that whether the question is fact based or description based. The questions that demand the exact answer are fact based whereas the questions that demands opinion are description based. This part is very useful in deciding what kind of response should be generated.

VI. CONCLUSION

In this paper we have studied the technique of automatic generation of expressive opinion in the response to the questions in the conversation. For that we have followed opinion making procedure where an opinion is generated with help of mining of opinion , sentence style conversion, sentence ranking methods. The input text is interpreted in factoid or non factoid questions. For that natural language processing, factoid sentence generation techniques are used.

Because of this there will be a development in conversation as it adds opinion in the response as well as expression to invite other person to take part in conversation with interest.

REFERENCES

[1] H. P. Grice et al, “Logic and conversation,” in “*Syntax and Semantics*, Academic, 1975, pp. 41–58”.

[2] B. Malinowski, “The problem of meaning in primitive languages, *Language and literacy in social practice: A reader”*, pp. 1–10, 1994.

[3] K. P. Schneider, *“Small Talk: Analysing Phatic Discourse*. Marburg,Germany: Hitzeroth”, 1988, vol. 1.

[4] P. D. Turney, “Thumbs up or thumbs down?: Semantic orientation applied
to unsupervised classification of reviews,” in *Proc. 40th Annu. Meeting Assoc. Comput. Linguist.*, 2002, pp. 417–424, ACL

[5] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: Sentiment classification using machine learning techniques,” in “*Proc. ACL-02 Conf. Empir. Meth. Natural Lang. Process.*, 2002, vol. 10, pp. 79–86, Association for Computational Linguistics”.

[6] T. Nakagawa, T. Kawada, K. Inui, and S. Kurohashi, "Extracting subjective and objective evaluative expressions from the web," in "*Proc. IEEE 2nd Int. Symp. Universal Commun. (ISUC'08)*, 2008, pp. 251–258".

[7] T. Kobayashi and S. Fujie, "Conversational robots: An approach to conversation protocol issues that utilizes the paralinguistic information available in a robot-human setting, *Acoust. Sci. Technol.*, vol. 34, no. 2, pp. 64–72, 2013".

[8] J. Kamps, M. Marx, R. J. Mokken, and M. De Rijke, "Using WordNet to measure semantic orientations of adjectives," in "*Proc. 4th Int. Conf. Lang. Resources Eval.*, 2004".

[9] C. Fellbaum, "WordNet: An electronic lexical database. 1998,2010[Online].Available:http://wwwcogsci.princeton.edu/wn"

[10] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in "*Proc. 2nd Int. Conf. Knowl. Capture*, 2003, 2004, pp. 70–77, ACM".

[11] Yoichi Matsuyama et al, "Automatic Expressive Opinion Sentence Generation" in 2015.