DIALOGUE ACT RECOGNITION FOR CONVERSATIONAL BOTS USING POS-TRIPLET FEATURES

SaiTeja.D[#1] , Harish.M[#2] , Abhishek Nair.M [#3] and G.Indra[*4]

[#]*UG Students, Dept. of CSE, R.M.K. College of Engineering and Technology, Chennai, India*
[*] *Asst. Prof., Dept. of CSE, R.M.K. College of Engineering and Technology, Chennai, India*

Abstract— **Conversational agents are the hot topic in the computer science field now. The main aim is to have natural conversations indistinguishable from the humans using algorithms from machine learning and natural language processing. Our project aims at a very important piece of this conversational agent's namely Dialogue recognition. A model has been proposed to identify the dialogue given by the user by training a machine learning model and also a comparison study is provided between two popular machine learning classifiers.**

Index Terms— Conversational agents, Natural language processing , Dialogue Recognition, Machine Learning.

I. INTRODUCTION

More and more companies are now resorting to the use of Chat bots to serve their customers. These Chatbots are specific to the domain of the companies that employ them. The main challenge of open domain conversational agents is to generate human like responses based on the entries that are given by the user and to guide the conversations in a meaningful way. Although there exists multiple algorithms that can train a machine learning model to identify and generate responses. There still exists some anomalies and the system may not be accurate at all the time. Modern Chatbots like Amazon's Echo , Apple's Siri make use of machine learning to provide advanced information retrieval processes in which responses are based on analysis of the results of the web search, while some other Chatbots like 'Mitsuku' makes use of pattern matching techniques to identify the user intent and generate appropriate responses.

II. DIALOGUE ACT RECOGNITION

The main aim of this subset in the design of Conversational agents is to identify the user intent. This means to find out what the user has to say or convey in the piece of text entered. Thus, this boils down to determine the function of the text/sentence of whether this is a question or a suggestion or an offer etc. The most common way of identifying this function is to build a statistical machine learning model based on some features extracted from the corpus. This model takes in the sentence as input and outputs the function that entry.

III. EXISTING MODELS

A. Bag of words

This model tends to ignore the structure, order, syntax of the sentences and basically counts the occurrence of each word in the entire corpus. This will generate a vocabulary of all the existing words in the corpus with their counts. Each of these word counts are counted as a feature. The intuition is that some words will be commonly occurring in some specific class, For e: wh- words (ie: what, where...) most commonly occur in the class"Question" .The major disadvantage of this model being that no importance is given to the word –orderings in the sentence, as there can exist two sentences with exactly the same words but different meanings whole-together.

B. Pattern matching

Sentences/text is treated as regular expressions and can be matched against the set of sentences that are labelled and exists in the agent's knowledge base. This technique not only is used for dialogue generation, but also for response generation which is again another important subset of the Conversational agent design. This is basically done using regular expressions and wildcard characters. This technique is used by the famous chatbot 'Mitsuku'. namely AIML in order to classify the sentences and generate required responses. The major drawback of this model is that it cannot infer anything on its own as for each and every sentence it has to search the entire knowledge-base, for finding the required pattern, as the size of knowledge-base increases the entire search process becomes an expensive approach.

IV. PROPOSED APPROACH

Here we are proposing a model that takes in the structure of the sentence/text given, and finds out the class to which the sentence/text belongs to. We are again using a machine learning model to train the classifier and making a comparison study between two most commonly used classifiers for dialogue recognition. This approach has been shown in figure-1.these entire process is divided into two phases: The feature extraction phase and the classification phase The model has been explained in section and then the corpus used and each phase is described respectively in the section below.
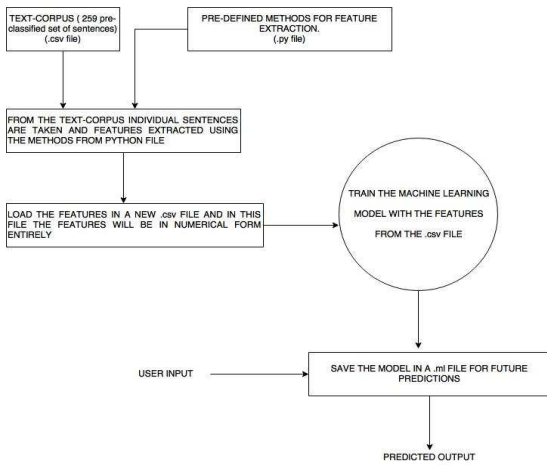
Figure-1

A. The pos-triplet model

The text corpus consists of the sentences that have been labelled with all the possible sentence namely, question, thanking, opening, closing, appreciation, statements. For each of these statements they are pre-processed by removing the stop words and lemmatising the words whenever necessary. This sentence is then converted to the pos tags and these pos tags are broken down into consecutive triplets maintaining the structure of the sentence intact, thus differing from the bag of word model. We then find out the set of most frequently occurring triplets in each class of the sentence and store it in a list for further prediction purposes. If these triplets occur in any of the sentence classes that feature class is incremented by one. This process is repeated for all the sentences in the corpus and the final feature vector is generated. At final this feature vector is given to the supervised machine learning model that can be used further to predict the class of the text/sentence entered by the user.

B. Corpus

The parent corpus that is being used is the switchboard corpus with the SWBD-DAMSL tag set. The corpus consists of 1155 5 min-telephone conversations which are further having 42 different dialogue types. But since we have classified the sentence to only six classes, rather than using the entire corpus, utterances matching to our classes are taken from this corpus and a subsidiary corpus for our model is used containing about 300 sentences classified into six separate classes.

C. Features used

There are a total of 17 features that are extracted from each of the sentence in the corpus. These basic features include:

a. The scores for each of the class that is obtained by comparing the pos-triplets, namely qscore (for questions statements), tscore (for thanking statements) and so on for all the six classes.

b. The no of question marks that exists in the statement and if it exists it gives a high score for the question class and adds extra weight to qscore feature.

c. The count of the words in the original sentence and the count of the words in the stemmed sentence (ie: after pre-processing of the statement).

d. The set of pos tags such as NN, NNP, CD, PRP that is Nouns, Proper Nouns, Count-words (two, two-hundred, etc.) and pronouns respectively each acting as an own separate feature.

e. The id for each statement is a unique key for identifying each sentence in the corpus uniquely this is computed by calculating the hash value of the sentence.

D. Classification

The feature vector is given to a machine learning classifier and initially Random Forest Classifier is being used to train the model, this classifier is chosen because prediction involves Multiple classes and the size of the dataset being comparatively smaller. The same dataset was also trained with the most popular machine learning these classifiers are noted in the below section.

V. RESULTS

After training the dataset, the most commonly occurring triplets for each of the sentence classes were noted as below in Table. 1.

SENTENCE CLASSES	POS TRIPLETS
Question	WP-VBP-PRP
Appreciation	NNS-RB-JJ
Thanking	NNP-PRP-IN
Opening	NNP-NN-NN
Closing	VBG-TO-PRP
Statements	VBG-NN-PRP

TABLE .1

A. Dialogue act classification

We observed the most frequently occurring pos – triplet from the training dataset given to the supervised machine learning model, and since because the training dataset consisted of only 300 classified sentences the frequency of the these triplets were considerably small. According to these triplets the features were extracted from the corpus and applied to the classifier. By using the random forest classifier model, after training the model, the model gave an accuracy of 94.5%.this was because of the considerably low size of the dataset, but the predictions were not up to the mark. There were wrong predictions made for many of the sentences given to the model, and this was probably because of the unevenness' of the sentence classes given in the dataset (ie:more sentences were given for class "Appreciation" as

compared to other classes).but a descent classifier was modelled using the random forest classifier.

B. Comparison with svm classifier

The same dataset was classified with the linear svm classifier and with the same dataset the accuracy score achieved was 84.9%.here also it was observed that due to the un evenness' of the classes in the corpus, the sentence class "Appreciation" was more predominant .this model gave a lower accuracy score compared to the random forest model and this is mainly due to size of the corpus and if the corpus would have been larger with more utterances, both these

Models would have predicted in almost the same manner.

VI. CONCLUSION

In this paper, we presented a new model for feature extraction for the dialogue act recognition problem using the pos-triplets against the known models such as bag of words and pattern matching. We trained the machine learning classifier with two known classifiers namely Random Forest model and the Support vector machines and achieved the accuracy score of 94% and 84.9% respectively. The model was trained against a hand-made corpus of text/sentences that was pre-classified into six separate classes. The main basic advantage of this model is that it preserves the structure of the sentence rather than just calculating the frequency of the occurrence of the words throughout the corpus. This rather makes the technique more useful as compared to its counterparts.

REFERENCES

[1] Broder Andrei, A taxonomy of web search, ACM SIGIR, V. 36, Issue 2, pp. 3-10, 2002.

[2] Cao Huanhuan, Hao Hu Derek, Shen Dou, Jiang Daxin, Sun Jian-Tao, Chen Enhong, Yang Qiang, Context-Aware Query Classification, The 32nd Annual ACM SIGIR Conference, pp. 3-10, 2009.

[3] Ganti Venkatesh, König Arnd Chistian, Li Xiao, Precomputing Search Features for Fast an Accurate Query Classification, Proceedings of the third ACM International Conference on Web Search and Data Mining ACM, pp. 61-70, 2010.

[4] Jansen Bernard J., Booth Danielle, Classifying Web Queries by Topic and User Intent, Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems, pp. 4285-4290, 2010.

[5] James Allen and Mark Core. 1997. Draft of DAMSL: Dialog Act Markup in Several Layers. Technical report, Multiparty Discourse Group. University of Rochester.

[6] Alexander Clark and Andrei Popescu-Belis. 2004. Multi-level Dialogue Act Tags. In Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue, pages 163–170. Association for Computational Linguistics.

[7] Daniel Jurafsky and James H. Martin. 2000. Speech and Language Processing: An Introduction to Natural Language Processing