An NLP algorithm to understand the action and environment of user in English language.

Saswata Das

Abstract – **Natural Language processing is a way of understanding and communicating with human speeches for any Artificial Intelligence. This project success in the extraction of subject, object, background and the actions mentioned in the human speech. The project uses parts of speech tagging, syntactic analysis and wordnet for the data extraction. The output of the project helps a Learning model to understand the human actions mentioned in their speech.**

Index Terms – **Parts of Speech Tagging, Natural Language Processing, Syntactic Analysis, WordNet.**

I. INTRODUCTION

Natural language processing was first developed in 1950 by Alan Turing. The system was of understanding human language for computer intelligence following a set of rules. Primarily it was used for parts of speech tagging. Then statistical machine translations were involved. Now the main goal of natural language processing is to extract valuable information from user speech, text or any form of input. These information is understood in various ways –

- Pattern evaluation from user input.
- Meaning evaluation from user input about what he or she is trying to make the AI understand.
- Sentiment analysis on various backgrounds like content quality understanding, content sentient understanding, content vulgarity checking etc.

Most common usages of the natural language processing are –

- Predicting user choices of enquiry by pattern analysis.
- Text to speech conversion.
- Understand the user mood by sentiment analysis.
- Understand user content effectiveness on others by sentiment analysis, vulgarity analysis etc.

Now today most NLP is concerned with parts of speech tagging. Mostly to understand what the user is saying and his actions. Common AI with speech tagging mechanisms are Google Now, Cortana, Siri etc. These AIs not just only predicts suggestions for user by their searches but can perform tasks given by the user. They understand what is the action to be performed and on what object it will be performed.

To feed the AI it mainly depends on the information rich words. The sentences used by the users are quite complex and have a lot of variations of formations. The AI must be able to understand the context of the sentence to figure out the user's actions. The speech structures being complex leads to ambiguity of understanding one's action.

But we know basic of grammar lies in the part of some subject doing some action n some object at a certain situation. Now if we can extract these four parameter any machine will be able to understand what the speech is about and the action that is mentioned. The speech gets simplified to a root level pattern of English language.

The program so developed targets mainly towards simplifying of the sentences to subject, object and background if any scenario present to understand the user actions. The program reads the sentences given by the user, passes it through various segments of the program, find the verbs and adjectives associated with the nouns of subject and object. Then finds out the background scene and print the parts. So to evaluate these verbs and adjectives for the AI to understand user.

II. METHODOLOGY

The program is developed using Python language to ease the internet accessibility and functionality. And the modified WordNet dictionary is made with Json. We select a story that is less complex for the program to read. The complexity is dependent on the factors like:

1. Grammar complexity
2. Words used
3. Relations made
4. Length of the sentence

The grammar complexities and relations of the sentences are resolved by decoding the grammar by sentence formation systems. They are of Simple, Compound, Complex and Compound-Complex. Words used are tracked by Wordnet system and using their simplified version. The length is manipulated by analyzing one sentence at a time internally.

The after taking the allowable input we start the program where asks for the input. As we provide the story it reads the Wordnet dictionary and takes the values into variables of arrays. Then story is broken into lines. Each line is processed at a single task. The words are extracted and analyzed one by one. Every word goes through various check process and manipulate process. If the word is found to be useful it is kept else removed. The words of subject or object are judged and take in consideration along with their related action.

The steps involved are –

1. Read the Wordnet dictionary.
2. Read each sentence make them to array of strings.
3. Analyse every array.

Manuscript received July, 2018.

Saswata Das*, Computer Science and Engineering , University of Engineering and Management, Kolkata, (dassaswata1996@gmail.com). Kolkata, West Bengal, India, Ph : 8420463112*

4. It tracks nouns, verbs, adjectives.
5. Shuffles them to make a sorted position of reading.
6. Removes unwanted words like articles, conjunctions etc.
7. Replaces the abstract words like pronouns with their actual word like noun.
8. Joins the adjectives and verbs to the related noun to make easier readability.
9. Finds background scenes, subjects and objects.
10. Makes a list, puts them in a json file and prints them.

III. OBSERVATION

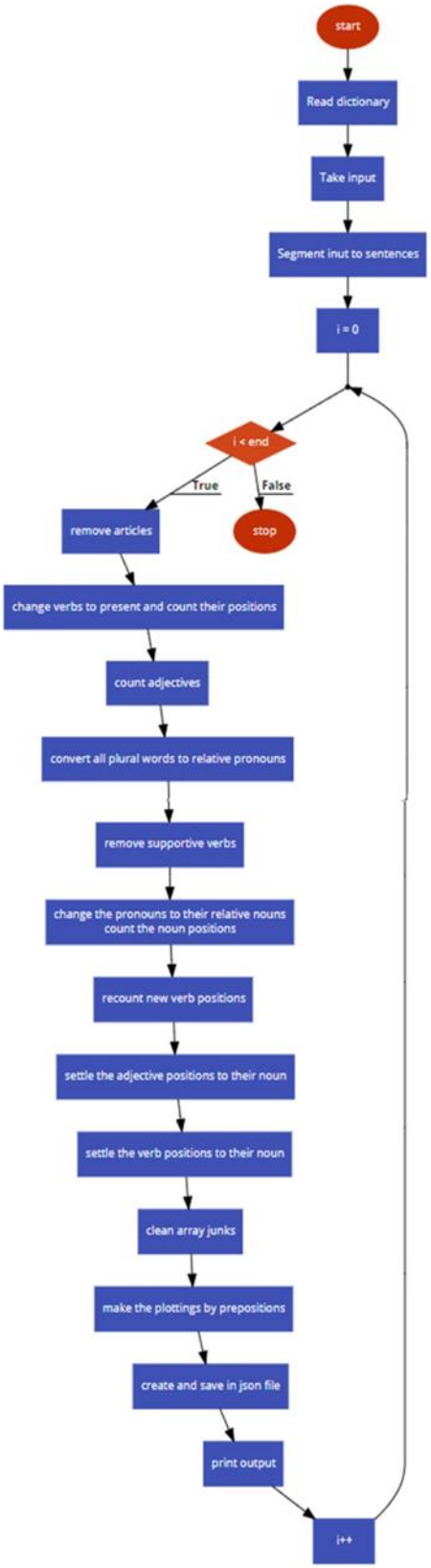
By running the above piece of code we will try to generate the subject, object and background for some input. We will provide the inputs at a simple structure to maintain scalability. The inputs are extracted from story pieces.
The result cases are (Table 1.1) :

Case no.	Input	Output
1	The heavy dog and cat chased the playing mouse. The mouse ran to the alley. The cat walked back to his park.	[{ "sub": "dog_chase-heavy|cat_chase", "obj": "mouse", "backg": "" }, { "sub": "mouse_run", "obj": "", "backg": "alley_back|" }, { "sub": "cat_walk", "obj": "", "backg": "park_back|" }]
2	The girl walked with the boy to the park.	[{ "sub": "girl_walk", "obj": "boy", "backg": "park_back|" }]
3	The boy walked by the dog on the street to the park. The boy then played with the girl.	[{ "sub": "boy_walk", "obj": "dog", "backg": "street_back|park_back|" }, { "sub": "boy_play", "obj": "girl", "backg": "" }]

(Table 1.1)

So we observe from the above tests that the program is able to extract subject , object and background. It also has verbs and adjectives related by joining them using special characters. The output is saved and showed in the form of Json texts.

IV. FLOWCHART



Fig(1.1)

V. RESULTS

After testing the program with various other inputs under both Unit and Integration testing methods we are quite sure about its functionality. The scalability is almost as expected though has a lot of scopes to be improved in near future. We also found that the program can automatically understands the verb relation and adjective relation to the nouns even though they might be complex in structure.
The results for heavy complex stories with too much of references to previous sentences are affecting the performance though but it still reaches the mark expected. With a periodic break in referential structure of sentences it is possible to get a full scale performance.

VI. CONCLUSION

After performing the above experiment with the program we find that it can successfully extract the required parameters like subject and object noun, verb and adjective related to it and the background situation noun.
The program has lot of scopes of improvement by hybridising it's functionality with prediction methods to ease the functional complexities. Learning methods can be used to improve the dictionary to improve understanding capacity.
The program has lot of future prospects like E-Learning, Story to video conversion, Deep Learning for story analysis and writing. Also the program can be developed to facilitate other program to ease their process of work.

ACKNOWLEDGMENT

I am thankful to International Journal of Advanced Research in Computer Engineering and Technology to provide me a platform to display my research. I would also be thankful for the easy and simple process of my paper presentation.

REFERENCES

[1] Prototype video *(Unpublished)*:
Extraction of Subject, Object and Background along with Action and Adjective.
https://drive.google.com/open?id=14tHngJ378IVkJEMBQGFa2Q_3WrfGSgkn
Developed/Uploaded : Saswata Das
[2] Wordnet:
What is WordNet? A conceptual Introduction with Python.
https://stevenloria.com/wordnet-tutorial/
Author : Steven Loria
[3] Books:
1. Natural Language Processing with Python
Author - Edward Loper, Ewan Klein, and Steven Bird.
2. An introduction to syntactic analysis an theory
Author - Dominique Sportiche, Hilda Koopman, Edward Stabler.
[4] Electronic Sources:
1. Sentence Structure:
https://learnenglish.britishcouncil.org/en/english-grammar/sentence-structure
Author : British Council
2. Sentence Structure:
https://www.english-grammar-revolution.com/sentence-structure.html
Author : Elizabeth O'Brien

Saswata Das
3rd Runner Up of Global Conference of Cyber Security 2017 Peach-a-thon Innovation Challenge and warded by honorable PM Sri. Narendra Modi.