IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.11431

Japanese Language Translator

Pranav Wadodkar[1], Bhumika Narkhede[2], Shivani Mahajan[3], Purva Patil[4]

U.G. Student, Department of Computer Engineering, SSBT's College of Engineering and Technology, Bambhori, Jalgaon, India[1-4]

Abstract: Language Translation is the process of detecting the language from any kind of Text, Text File, or an Image and then translate it into target language. Implementation of a Natural Language Processing model for Language Translation is to be carried out in this project. The major task is to identify those features or parameters which could be used to clearly distinguish the language and translate it. This model makes use of Machine Learning, Artificial Intelligence, and Natural Language Processing. The project aims at detecting Japanese and translate it into English. Experiments were conducted by forming text samples obtained from online articles and social media. This corpus comprises of general articles, each of them spanning over at least 100 words. The entire corpus is split into two sets, larger unit as the training data-set and a smaller set as the test set.

Keywords: Language, Translation, Japanese, English.

I. INTRODUCTION

The phenomenon of globalization has brought together people from around the world. However, one barrier to this increase in global communication is that many people communicate in different languages and effectively lack a common communication medium. That is, in order to communicate effectively a language which is mutually understandable by both parties is required. Language Translation offers a means for providing this medium. Language as a communication system is thought to be fundamentally different from and of much higher complexity than those of other species as it is based on a complex system of rules relating symbols to their meanings, resulting in an indefinite number of possible innovative utterances from a finite number of elements [1]. Among the various factors that define different cultures and communities, an important factor is language. The importance of language for human to human communication can be over emphasized. Language Translation is an important task for information retrieval services. This paper presents the implementation of a platform for language translation in multi-lingual documents on web. The platform consists of modules to achieve the tasks automatically. Furthermore, artificial neural networks were used for the translation of languages in multi-lingual documents. Results for six languages including English, French, Spanish, Italian, Russian, are present. The major benefit of the approach is that the ANN based language translation system could meet the expectations in real-time language translate accuracy with the help of a developed system.

II. ANALYSIS

A. Hardware Requirement and Software Requirements

Requirement analysis help the software engineer to better understand the problem they will work to solve. It includes that set of tasks that lead to an understanding: -

Server side:

- Intel core due 3, 5, 7 processers.
- Window server environment with python Requirement.
- 2 gigabyte of random access memory.

Clint side: A user should browser to access the language translation services. Server side: Server should have support vector support machine library to identify language using NLP.

B. Functional Requirements

- The service expected of system to translate language the given input, the system recognizes the language irrespective the type of given input.
- Given text sample system should extract NLP and detect the language.
- Input file sample or image should not be containing any garbage.

C. Non Functional Requirements

- The user is Expected to have stable internet connection.
- The User is expected to Provide input file.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.11431

- The input file should be of specified format.

III. DESIGN

The proposed system architecture is show fig. The NLP system takes text sample as the input. There are three processing blocks in the architecture; they are, pre-processing blocks, feature Extraction blocks and machine learning block. This cepstral feature vector serves as the input to the next block which is the Machine Learning Block. The SVM in the training phase creates a model based on the input feature vectors for different languages. This model file is used by the classifier in the testing phase to predict language [2]. The final result is the language identified for the given test samples. A high level design of the system and it follows client server architecture. The client side comprise of a portal which allows the user to upload a text sample using a browser [3]. The sample is sent over the network to the remote server running an NLP system.

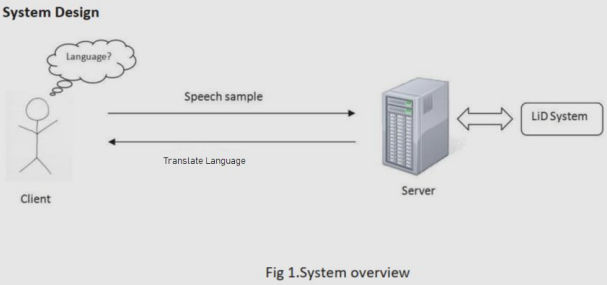

Fig 1.System overview

IV. ALGORITHM

Input: Set of Spellcheckers, Li; unknown document, D; Benchmark Specification, BM.
Output: Translation of the input document.
Begin:

1. Pre-process unknown document and tokenize into words
2. Remove all numeric words and all special characters
3. Convert all words into lowercase and sort words by word length
4. Index word list into set such that each word is searched only once
5. for each word wD
6. if length (w) = n
7. for language in Li
8. if w in lang-word-length (n)
9. lang-word-count = lang-word-count + 1 increment word count
10. end if
11. end for
12. end if
13. end for
14. Compute matrix totals
15. for language in Li
16. if percentage (%) of the highest scoring Spellchecker BM(BenchMark)
17. The document is translated
18. else document is untranslated
19. end for
20. End

V. RESULT

After training the dataset on three different machine learning model the outcome that has been extracted is as follows, the linear regression model performed the best with the score of approximately 91%, followed by the lasso regression with an approximate score of 91%, lastly the decision tree score almost 90% when trained on dataset. As the final decision to choose over these machine learning models the optimal choice had to be the linear regression model and hence, that is the reason of using the same in the proposed solution.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.11431

True Positive 260	False Positive 19
True Negative 20	False Negative 135

VI. CONCLUSION

The current system is capable of translate Japanese language to English Successfully. The system can distinguish between Japanese and English words. The translation accuracy is approximately 91%.

REFERENCES

[1]. Miyako Inoue. Speech without a speaking body: "Japanese women's language" in translation. Language & Communication, 23(3-4):315-330, 2003.
[2]. Makoto Nagao. A framework of a mechanical translation between Japanese and English by analogy principle. Na, 1984.
[3]. Robert Weschler. Uses pf Japanese in the engligh classroom: Introducing the functional translation method. Kyoritsu Women's University Department of International Studies Journal, 12:87-110, 1997.