LANGUAGE TRANSLATION USING AI/ML

ADITYA BHAMBRE, NAYAN ZAGADE ,CHINMAY KALE

Students, Department of Computer Engineering, D.Y Patil School OF Engineering &Technology, Pune Ambi, Maharashtra,India

---

Abstract - *A language translator is an application that can be used to translate from one dialect to another. There have been difficulties in information communication between countries over the years. In modern times, language interpreters must understand and speak boththe language that is being translated and translated. The aim is to develop an android language converter app to make learning and language translation easy and facilitate stress-free communication, which can work more efficiently than other existing applicationswithanoptimized code for thetranslation process. The motto is to implement a lens for scanning the images instantly converting them to text and translating them. Making the app work even in offline mode is the main objective.*

***Key Words*: Language, Translation, Communication, Lens, Offline**

1. INTRODUCTION

A Translator is an assistant for translation that helps people survive in places where the language is unknown. They can easily input their message via text or speech and translate it into any language. This can also assist people with difficulty in writing with the help of speech recognition features. The text in an image canbe captured with a lens and it can be translated into any language as per the user's choice. I have used specific codes that process two or more combined user input modes, such as speech and image. Even for detecting the voice and for scanning images we used Java code to perform the actions, unlike other systems. In our proposed model itworks in both online and offline mode. The proposed model is the structure of the next generation of voice Assistants.

2. LITERATURE SURVEY

This chapter will discuss an article review of our project. The article review is important because it is used to help the developer build the system so that the developer gets more knowledge of the pros and cons of the system, whichhelps the developer choose the best way to develop the system. Research and analysis of an existing system or current system have to be done to build a good system. A good system always comes after the enhancement of existing systems.

A Practical Guide for Translators, 5th ed, Bristol, Multilingual Matters by Samuelsson-Brown G in 2020 briefs about translators and their applications which helped to buildthe flow of developing the app effectively

Building a translation competence model by Amsterdam, John Benjamins in 2021 gave guidance on building the model

Triangulating Translation: Perspectives in Process Oriented Research by Amsterdam, John Benjamins in 2021 says about translation and its applications which helped in theimplementation of translation more efficiently

Approaches to Translation by Newmark P in 1981 tells how translation works efficiently and guided to complete the app successfully

The research helped in the completion of the applicationand served as a stepping stone for achieving efficiency in translation of text to text, speech-to-text, and text translation.

3. PROBLEM STATEMENT

Communicating with people at one time when the user doesn't understand the language the other user speaks. By using this app, the user can easily understand whathe/she (another user) is speaking and translate it to the userwho uses the app. They are completely useful for the one who travels from country to country or from state

to state i.e. the one who travels the world. There are many apps like this but the app which has been developed can work completely in offline mode. Moreover, it can recognize the text in an image accurately and translate it to the language as per the user's choice which reduces the efforts of the user in typing the text present in anything again in a translator app.

4. REQUIREMENTS

Tools and Resources used –

4.1 Jupyter Notebook on Kaggle
4.2 French to English Dataset
4.3 Keras as the main library
4.4 Pandas for pre-processing
4.5 Natural Language Toolkit
4.6 Matplotlib for graphical representations

4.7 HARDWARE REQUIREMENTS

S.NO	REQUIREMENT	HARDWARE
1	Processor	Intel Pentium III
2	RAM	minimum 4 GB
3	HDD	40 GB
4	Secondary Storage	1.44 MB FDD, CD-R, CD+RW CD
5	Monitor	14” Color Monitor

4.8 FUNCTIONAL REQUIREMENTS

1. System Admin: Installs translations in the application.

2. Translator: Updates translations that can then be used in the application.

3. Developer: Creates windows and customizations that use a translation

framework to enable a language-specific version of the customization.

4. User: Works with the functionality in the application and specifies the language

they would like to view the user interface. Chooses a language from those that

are active within the application.

5. FLOW CHART

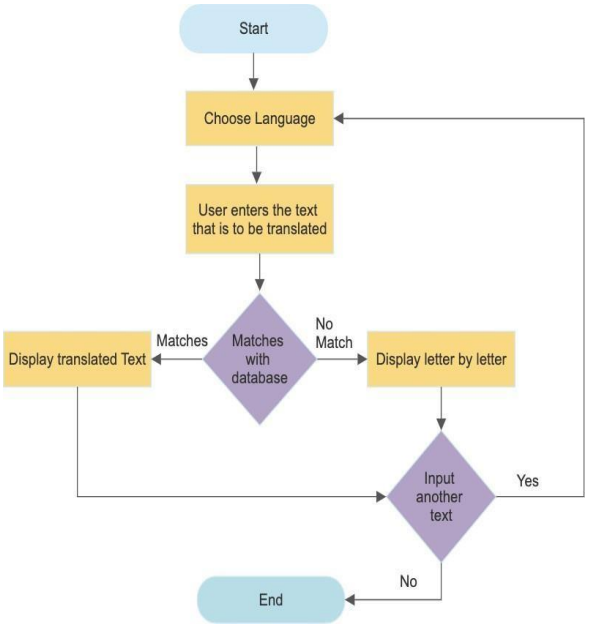


Fig 1: Flow of the application

6. METHODOLOGY

6.1 LSTM basic definition:

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture designed to overcome the limitations of standard RNNs in capturing and retaining long-range dependencies in sequential data. LSTMs are particularly well-suited for tasks involving time series data, natural language processing, and other sequential data analysis.

6.2 RNNs and the Need for LSTMs:

Standard RNNs process sequential data by maintaining a hidden state that evolves as it receives new input elements in a sequence. However, they suffer from the "vanishing gradient" problem, which makes it challenging for them to learn long-term dependencies. This is because gradients that become too small during backpropagation prevent earlier time steps from effectively influencing later ones.

6.3 LSTM Introduction:

LSTMs were introduced to address this problem by introducing a more complex structure within the recurrent unit, which allows them to better capture and propagate information across long sequences.

7. RESULT SCREENSHOTS

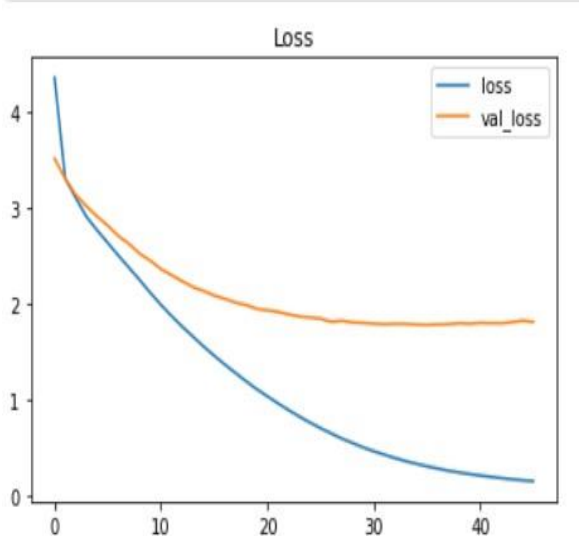
Result on the Training Set

FRENCH (SOURCE) IN ENGLISH	ENGLISH (TARGET)	AUTOMATIC TRANSLATION
nous en savons assez	we know enough	we know enough
garde ton sang froid	stay calm	stay calm
je ne pleurerai pas	i won t cry	i won t cry
je ne suis pas contente	i m not happy	i m not happy
moi je veux ça	i want that	i want that
j étais tellement heureuse	i was so happy	i was so happy
j aime le printemps	i like spring	i like spring
c est mon garçon	that s my boy	that s my dog
mille mercis	many thanks	thanks a bunch
quelle horreur	how horrible	how horrible
soyez satisfaites	be content	be content
toi décide	you decide	you decide
je m en suis remis	i recovered	i recovered
ce sont les affaires	it s business	it s business
je dois m en aller	i need to go	i must to go
sommes nous prêtes	are we ready	are we ready
arrêtez de crier	stop shouting	stop shouting
je lis souvent	i often read	i often read
les plantes croissent	plants grow	plants grow
il m a fallu le faire	i had to do it	i had to do it
nous éclatâmes de rire	we broke up	we broke up

Result on the Test Set

FRENCH (SOURCE) IN ENGLISH	ENGLISH (TARGET)	AUTOMATIC TRANSLATION
ils ont abandonné	they gave up	they gave up
rappelle moi	call me back	help me
je veux essayer	i want to try	i want to try
ça fonctionne bien	it works well	it works well
grimpe dans la camionnette	get in the van	get in the van
je suis mince	i m thin	i m stunned
elle semble riche	she seems rich	she is rich
ça me gave	this annoys me	it resent happen
c était long	it was long	that worked
c était un mensonge	it was a lie	that was a lie
conduis toi en homme	act like a man	get your luck
laissez moi m en occuper	leave it to me	let me sleep
puis je manger ceci	may i eat this	can i eat this
devine	make a guess	turn it
je ne suis pas jolie	i m not pretty	i m not hard
demande à quiconque	ask anyone	ask security
venez nous rejoindre	come join us	come us
vous ennuyez vous	are you bored	you you stay
je ne viendrai pas	i won t come	i can t swim
c est un voleur	he is a thief	it s a fake
bien joué	well done	good job

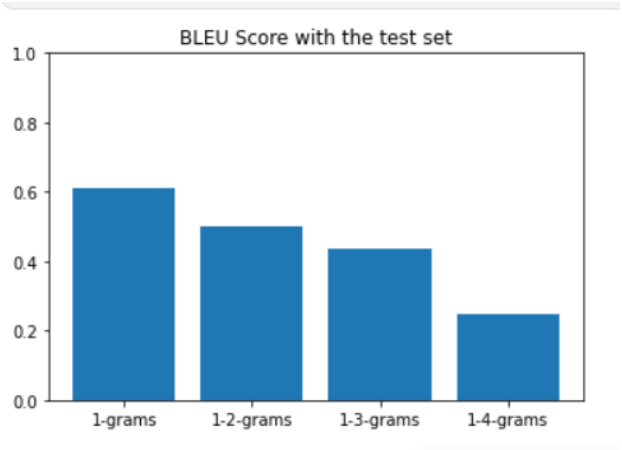
:

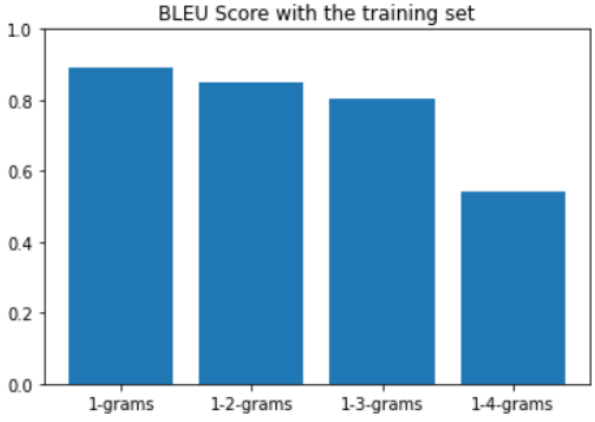


Methodology for training

The methodology to process, train, and test the LSTM model isas follows –

1. Data Pre-processing – Here we use Pandas to performpre-processing of the CSV file data i.e. they create the train and test sets for further use.
2. Create and train model – Here the Kera's and NLTKlibraries for ML/AI are used to build and train the model. It includes models like LSTM, ‘corpus_belu’, Tokenizer, etc.
3. Result on test set – Towards the end a portion of thedataset will be used to test the model.
4. Prediction evaluation with BELU – Also BELU algorithm will be used to further investigate the accuracy of the LSTM model, here we use NLTK’s‘corpus_belu’ method.
5. The result of the BELU score on training and test sets.





8. Conclusion

- Deliver information in multiple languages
- Improve communication in multiple languages
- Increase human translation productivity
- Create and manage enterprise language as a corporate asset
- Integrate with enterprise application
- Assist people with disability in speech
- Capture images and translate them easily

Further future work can be taken up on language Recognition for text which is handwritten and also includes an option for uploading a document for translation

9. REFERENCES

[1] 1Sim Liew Fong Faculty of Information Sciences and Engineering, Management and Science University, Shah Alam, Selangor Darul Ehsan, Malaysia(chapter 8)

[2] Scientific World Journal Volume 2014 (2020), Article ID 267872

[3] Md Gapar bin Md JoharFaculty of Information Sciences and Engineering, Management, and Science University, ShahAlam, Selangor Darul Ehsan, Malaysia(chapter-8)

[4] Baker, M. (ed.) (2020) Routledge Encyclopedia of Translation Studies, London, Routledge.

[5] Newmark, P. (1981) Approaches to Translation, Oxford, Pergamon.

[6] Encyclopedia of Literary Translation into English by Olive Classe (Editor).