



Bridging the Language Gap: NLP Based Translation of Government Schemes for Marathi-Speaking Farmers

Pooja Suryabhan Gaygaye¹, Ms.Yogita Kadbane²

Savitribai Phule Pune University,Pune,Maharashtra,India^{1,2}

Abstract: Government provide many schemes for farmer.the scheme is on official portal,but most of the scheme are in English .In Maharashtra many Farmers know only Marathi language,they did not aware about English language,due to this they are not able to take the benefits of the governments schemes.

My Research objectives is ,all farmers can get benefited from governments schemes,they can get advantage of the government policies also. I am used Experimental Methodology and I am used A Transformer-based Neural Machine Translation .the performance of model was evaluated using BLEU and CHRF metrics.the system helps to farmers for easily understand government scheme information,specially for Marathi speaking farmers.In future I am focusing on expanding the Data set of farmer related schemes,supporting additional Indian languages and voice based interaction.

Keyword: Natural Language Processing(NLP),English-Marathi Translation ,Transformer,Neural Machine Translation,Government Schemes,Agricultural Information,Marathi Language

INTRODUCTION

Indian governments introduce many schemes to farmers ,that scheme is related with subsidy,insurance,flower crops,fruit crops,irrigation,loans,etc.this scheme are published on government official website like MAHADBT portal or pm KISAN etc.in Maharashtra many farmers speak Marathi,they are not able to understand this English language or Hindi language scheme,they do not take the benefits of government scheme,they get left out form this.due to this,language play an important role ,why farmers is not taking the benefit of government schemes.All information related scheme is available on portal like eligibility,documents,deadlines,but farmers are not able to Understand because lack of literacy.when we go for Manual Translation it becomes time consuming ,and it may contain sometime error.

Due to above problem there is need of English to Marathi translation system,by using which the farmer can acquire the benefits of all scheme.

For this I am using natural language processing ,NLP is a branch of AI.NLP automatically translates the text of one language to another language using Tokenizers.so,there is a need of NLP based system.

English and Marathi are two opposite sides of one mirror.English follows subjectverb-object word order while the Marathi follows subject-object-verb word order.in this I am translating the word by word ,a good translation system must preserve the meaning,terminology,structure,idea,etc.

LITERATURE REVIEW

In Previous studies I have explored that,machine learning and natural language processing different techniques can help to translate the sentences form one language to another language.

In this different people used different technologies such as,some used Statistical

Machine Translation and Seq2Seq models and other used LSTM, and Transformer models at the end transformer based model get high success because they understand the relationship between the English word and its similar Marathi words.Some Researchers used NLP for agriculture area to provide the knowledge about crops,crops related diseases,Irrigation,etc.This proves that NLP can be useful for delivering information.

Previous researchers also provide ,government scheme information through AI Chat bots,multilingual systems,etc.but they are in formal language,so it is difficult for farmers to understand.from literature review I have understand that,several studies have worked on English to Marathi Translation,NLP based agriculture applications and multilingual information system.but there is limited research in this area. I found the research gap that,the translation system exists,but they are not in Marathi language.the system are in the form of chat bots or in app format.they are not focusing specifically government schemes,or schemes related to farmers,so I found this research gap.



METHODOLOGY

This research follows experimental research. experimental research is a process in which ,we do something but the variables are in our control.this research studies existing English to Marathi translation method and create a machine translation system which is only designed for farmer and and government scheme.

For this system ,I am collected data from government portals,and then converted that data into English to Marathi sentences pairs.that data set contains information related to agricultural scheme,subsidies,Irrigation,etc.I am clean the data by using libraries.in this firstly I am remove duplicates,remove unnecessary spaces,handles missing values,correct incorrect text,etc.I am divided the data set into 3 parts,I.e. training dataused for training the model,testing data-which is used for testing model and validation data-which is used for check and improve the model training. A transformer based neural machine is used for translation of English to Marathi government scheme,in this the model learns the relationship between the English sentences with respect to its Marathi sentences.I am train the model by using English to Marathi parallel data set. Translation quality of model is measure using BLEU score and chrF score and also human evaluation is included.

RESULTS

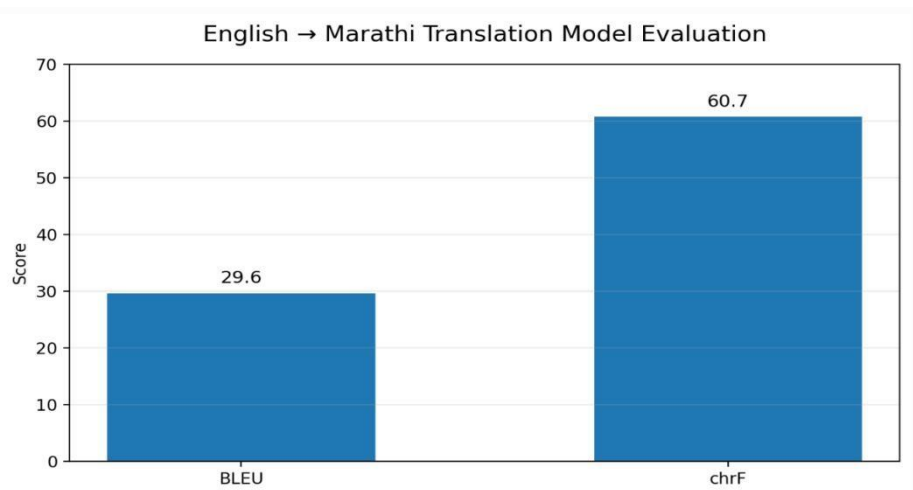
In this I am used the data set which included the English and Marathi sentences pair which is related to government schemes and farmer related data.I am divided the data set into training data,testing data and validation data. The transformer based model is trained using the prepared parallel data set.in training the model learned the relationship between the input sentences and output sentences.the loss of training is decreased during training process,stating that ,the model is learned from the data set.

Following is my some translation results:

Translation Examples: English to Marathi

English Input	Reference Marathi	Model Output
Farmers can apply for the scheme online.	शेतकरी या योजनेसाठी ऑनलाईन अर्ज करू शकतात.	शेतकरी या योजनेसाठी ऑनलाईन अर्ज करू शकतात.
The scheme provides financial assistance to farmers.	ही योजना शेतकऱ्यांना आर्थिक मदत प्रदान करते.	ही योजना शेतकऱ्यांना आर्थिक मदत प्रदान करते.

The model can translate several English sentences which is related to farmer and government scheme into Marathi.in this some sentences can contained grammatical mistakes or incorrect word choices,particularly for domain specific terms.



I am using BLEU score to measure the similarity between the generated Marathi translation and the reference translation.and also using CHRF score to calculated character level similarity and it can be useful for morphological rich languages such as Marathi.

During evaluation ,there is some limitations are observed.the model sometimes produce incorrect word choices and sometimes it take a little extra time for processing,because it find the more accurate word for the Marathi word form the data set.



DISCUSSION

The developed model can translate the English government scheme sentences into Marathi. the model was able to translate the understandable sentences from English to Marathi for farmer used. for Evaluation I am used BLEU and CHRF, this score indicates the level of similarity between the generated Marathi translation and the reference translation. the score, which is get from the evaluation, that state that model is able to learn common English to Marathi translation patterns. In this I am observed some errors like incorrect word choices and it take a little extra time for giving the output. it is because the model check the English word with 5 Marathi word and choose the most appropriate word from the data set. This finding is consistent as compare to previous research. previous research says that transformer based approaches can be effectively used for neural machine translation. however the present study focuses specifically on government scheme information for farmers, where domain specific vocabulary and simple Marathi are important.

The system has the potential to make the government scheme information more accessible to Marathi speaking farmers. when we provide information into simple Marathi, it may help to understand scheme benefits, eligibility and application process more easily.

CONCLUSION

The reserch is focusing on developing an NLP based system, which can translates the governments scheme which is related to farmer or agriculture, into the regional language i.e. Marathi language. for building this system I am used a english to marathi parallel dataset, which contains the sentences from government official website. Evaluation is done using BLEU and CHRF metrics. this both metrics help me to measure the quality of measured sentences. the results show that the model may face the difficulties with complex sentences or its speed may get slow. When we provide the information in regional language, then many farmers get benefited from this.

Limitation:

In future the data set is in small size with specific government schemes and not each and every government scheme are included in it. The system can also be improved by including more domain specific vocabulary, better translation models, voice input and output, etc. This improvements can help develop a more useful multilingual information system for farmers.

REFERENCES

- [1]. P. Salunke, A. Kadam, S. Joshi, et al., "Hybrid machine translation for English to Marathi: A research evaluation in Machine Translation: (Hybrid translator)", 2016
- [2]. S. Shilaskar, S. Ghadge, M. Dhopade, et al., "English to Marathi Text Translation using Deep Learning", 2022
- [3]. P. Pandit, A. Jangale, et al., "Saajhedar multilingual agricultural AI scheme Assistant", 2026
- [4]. Y. Jhaveri, A. Mahajan, C. Bhadane, et al., "Neural Machine Translation from English to Marathi using various Techniques", 2023
- [5]. M. Bilal, G. G., "AgriGov: A Structured Multilingual Dataset Curation for Indian Government Schemes for Farmers", 2026
- [6]. B. Mehra, J. Anitha, "Empowering Indian Farmers with Multilingual AI", 2025
- [7]. A. Godase, S. Govilkar, et al., "Machine Translation development for indian languages and its approaches", 2015
- [8]. S. Khanvilkar, S. Beatrice, et al., "English to Marathi translation: A survey of Parallel corpora, datasets, Methodologies and Tools", 2025
- [9]. N. Telrandhe, R. kadu, "From standard Marathi to dialects: A review of machine translation techniques and future directions", 2025
- [10]. U. Patkar, "Machine Translation of English to Ahirani language: A Review", 2020
- [11]. U. Patkar, "An hybrid and Syntactic Machine Translation Model For English to Ahirani Language", 2021
- [12]. P. Lahoti, N. Mittal, et al., "A survey on nlp resources, tools, and techniques for
- [13]. Marathi language processing", 2022
- [14]. R. Raja, A. Vats, "Parallel corpora for machine translation in low resource indic languages: A comprehensive review", 2025
- [15]. Y. Jhaveri, A. Mahajan, et al., "Neural Machine Translation from English to Marathi using various Techniques", 2022
- [16]. S. Sani, S. Vijaya, et al., "A survey on the machine Translation methods for Indian languages: challenges, availability and production of parallel corpora, government policies and research", 2024