ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 11 N. 4 (2022), 411-425
eISSN: 2255-2863
DOI: https://doi.org/10.14201/adcaij.28380

A Novel Framework for Ancient Text Translation Using Artificial Intelligence

Dr. Shikha Chadha[1], Ms. Neha Gupta[2], Dr. Anil B C[3], and Ms. Rosey Chauhan[4]

[1, 2, 4] Department of Information Technology, JSS Academy of Technical Education, Noida
[3] Department of CSE(AIML), JSS Academy of Technical Education, Bengaluru
[c]Computer Sciences Department, University of Salamanca, 1 Escuelas St., Salamanca, 37003
shikhaverma@jssaten.ac.in, neha.gutta@jssaten.ac.in, anilbc@jssateb.ac.in, roseychauhan@jssaten.ac.in

KEYWORDS

ancient text, Artificial Intelligence (AI); Long Short Term Memory (LSTM); translation

ABSTRACT

Ancient scripts have been acting as repositories of knowledge, culture and of the history of civilization. To have greater access to the valuable information present in ancient scripts, an appropriate translation system needs to be developed to adapt to the complexity of the scripts and the lack of knowledge of the reader. In this study, a prediction and translation system has been implemented using Artificial Intelligence. The system has been trained using the Sunda-Dataset and a self-generated dataset. The translation of an ancient script, viz. from Sundanese to English, is done using a two layer Recurrent Neural Network. The technique used in this proposal is compared with a pre-existing translator called IM Translator. The results show that the BLEU score increased by 8% in comparison to IM Translator. Furthermore, WER decreased by 10% in contrast to IM Translator. Finally, the N-Gram analysis results indicate a 3% to 4% increase in 100% contrast value.

1. Introduction

Ancient scripts have been a great source of cultural and civilization knowledge with a lot of information of Vedas and Upanishads flowing in them that need to be preserved for future use. Most ancient scripts are available in degraded format and their complexity proves to be a great challenge for researchers who extract and translate their text (Chadha *et al.* 2015), (Wang *et al.*, 2016). To preserve them for future use, a system needs to be developed to translate scripts into a modern language.

Essentially, the paper focuses on a system for translating ancient languages into a universal language (Miyamoto, Y., & Cho, 2016), so that the valuable collection of information can flow from one generation to another.

The main motivation behind this study has been to give travelers from different parts of the world access to information about a rich and valuable culture which has the potential to increase tourism and ensures resources remain preserved for future generations. In contrast to it various scholars working on classical ancient scientific texts have a continuous tradition of discourse.

For the purpose of translating an ancient script into a universal language, Long Short Term Memory (LSTM) (Bijalwan *et al*., 2014), which is a technique based on RNN, has been used. As the LSTM does not require any linguistic rules, it further enables vector sequencing and uses many to many mapping to design the translator. It even solves the problem of gradient vanishing, as it stores the previous output and on its basis, the next state is given.

There are several translators that convert Sundanese text into English, such as IM Translator. net, Google Translator and stars21.com, which mostly use either a phrase-based statistical translator (PBMST) or Dictionary Based Translation, though the result of the dictionary-based translator is not very precise, as dictionary-based translators (Apriyanti *et al.* 2016), (Athiwaratkun & Stokes, 2016) translate the sentences word by word. Whereas PBMST (Bijalwan et al., 2014) translates the whole sentence using gradient vanishing, therefore meaning is preserved and the translation becomes more relevant (Chaudhary & Patel, 2018) (Jussà & Fonollosa, 2016).

The contribution of the paper is to translate the low resource Sundanese language into a universally accepted language i.e., English, being a low resource language pair translation, various sub-objectives are accomplished as stated.

- To collect ancient text using primary and secondary data collection techniques.
- To pre-process the ancient script using text cleansing and vectorization with one hot vector encoding in order to convert the input text into tensors.
- To generate a text translation and prediction model for translating the ancient script using LSTM into a current recognizable language i.e. English.
- To verify and validate (V&V) the developed ancient script text translation and prediction model.

1.1. Definition of Terms

- Vectorization: Is the process of converting a word into a large sequence of numbers that may hold the sequence of a complex structure, which can then be interpreted by a computer using various data mining and machine learning algorithms.
- Word Cleansing: Word cleansing includes the removal of stop words, such as punctuation marks, converting all uppercase letters into lower case letters which helps reduce the size of the source text to 75%.
- One Hot Encoding: One hot encoding is the process of Machine Learning (ML) that converts the data into some number of categorical value, which is fixed to a limited number of sets, with each value representing a different category. Thus, the input text characters are mapped on a numeric index so that each character has a unique index.

The rest of paper is organized as follows: Section 2 provides background on state-of-the-art research into various translation methods; Section 3 summarizes the data collection process of the Sundanese text, Section 4 contextualizes our models with respect to the experiment done using LSTM, Section 5 explores the obtained results and the analysis carried out on the Sundanese dataset, ending with a conclusion and a description of future lines of research.

1.2. Related Work

The technique presented in (Suryani et al. 2017) used Link Grammar (LG) and Statistical-Based Translation Technique (SBTT) to decide on the correct translation of the text. English prepositions can have more than one translation meaning in Indonesian text, so according to LG the word before the preposition is the word explained by it and the word after the preposition explains the preposition. Following the application of LG, the statistical method was used to translate the word. The Bilingual Evaluation Understudy (BLEU) showed that the translation of the preposition "with" in sentences achieved a precision of 81%, which is higher than that of Google's translator. The technique proposed in (Ray *et al.*, 2015) for the translation of a text from a low resource language to a target language with a limited parallel corpus involved making use of a rich resource language which was similar to the original source language, taking advantage of overlapping vocabulary. In the study, BLEU showed improvement in the sentence to 2-5 points. (Muzaini *et al.*, 2018) proposed a technique that focuses on clause identification and plays an important role in the extraction of complex sentences. The sentences were fed into a translator for translation and then passed to the parser tree. 2000 sentences had been used and BLUE score was found to be around 31.75%. The study proposed by (Hinton *et al.*, 2018) applied Part-Of-Speech Tagger (PoS Tagger) in a translator to get more efficient translation results. The Sundanese dataset was used for modeling and the results were better than if only surface form had been used, nevertheless, some problems were appeared during experiments, such as Out of Word Vocabulary (OOV) which was caused by a low amount of parallel corpus and noise in the ancient text. (Xiaoyuan *et al.*, 2017) proposed a study for comparison between Recurrent Neural Network (RNN) and statistical-based network with n-gram model in which English-Indonesian Machine Translation (MT) and vice versa were conducted. The results indicate BLEU score and Rank-based Intuitive Bilingual Evaluation Score (RIBES) increased by 1.1 and 1.6 higher than statistical-based technique. (Lauriola *et al.*, 2022) proposed a technique where clause identification plays an important role in the extraction of complex sentences. The sentences were fed into a translator and then passed to the parser tree. The BLEU score indicated that the resulting translation was at around 31.75%. (Chadha *et al.*, 2020) used Rule-Based Machine Translation (RBMT), the shift reducing parsing was used for linguistic information on the source language and PoS tagger was used for word class. A bilingual dictionary was used for translation and achieved an accuracy of 93.33%. (Zhou *et al.*, 2016) proposed a model for translating English to Indonesian language translation and vice versa using grammatical structure, cultural words, and writing mechanics. The calculated average accuracy was at 0.1163.

The study done by (Huang *et al.*, 2018) performed the translation of a sentence from one language to another which required a better understanding of the source and target languages. A hybrid approach was used for translation which combined example-based machine translation and transfer approaches that exhibit an accuracy of 75%. (Zhang *et al.*, 2018) proposed a Neural Machine Translation (NMT) system that used character-based embedding in combination in spite of word-based embedding. Convolution layers were used to replace the standard lookup-based word representations. BLEU points increased up to 3 points in the German-English translation task. (Lin & Shen, 2018) proposed an LSTM

based language model and a Gated Recurrent Unit (GRU) language model. It used an attention mechanism similar to (Cho *et al*., 2014) from the machine translation.

The study done by (Maitra *et al*., 2015) proposed a system based on RNN and Encoder-Decoder for generating quatrains taking keywords as an input. The system learns the semantic meaning in the sentence and learns semantic meaning among the sentences in the poem. (Nurseitov *et al.*, 2021) proposed a machine translation technique using deep learning, Tanaka corpus was used to convert the Japanese language into the English language. In the above approach neural machine translation was used which belonged to a family of encoder–decoders in which the encoder encodes a source sentence into a fixed-length vector called tensors from which a decoder generates a translation. (Ray *et al*., 2015) proposed a model using RNN-LSTM for translating Arabic text into English. COCO caption dataset was built, the performance of the proposed model on the test dataset gave a result of 46.2 for the BLEU-1 score.

In (Swe & Tin, 2021) a system was proposed for a segmentation free translation system using Long Short Term Memory (LSTM) and training was done on word and text line level. Precision Pattern Rate (PPR) and Recall Precision Rate (RPR) was calculated to be 80%. (Singh *et al*., 2017) proposed the use of RNN-CNN along with a Generative-Adversarial Model for text encoding to convert images into text. The dataset used was Oxford 102flowers, and there were 19 layers of CNN (Convolution Neural Network), achieving an efficiency of 23.4 according to BLEU. (Haroon & Shaharban, 2016) proposed a system based on end-to-end neural networks for converting ancient Chinese language into contemporary Chinese language. It achieved an F-score of 94.4% and a BLEU score of 26.95 for translation from ancient into contemporary language and a score of 36.34 BLEU when translating from contemporary into ancient.

(Windu *et al.*, 2016) focused on various tasks that demonstrated the application of deep learning techniques in Natural Language Processing, different hardware, software and popular corpora were used for this purpose. The multiple NLP classifications were used, such as sequence classification, pair wise classification and word labeling etc. The typical structure was used for sequence word classification, with static vectors used as input. Super GLUE and GLUE Benchmark values were calculated.

In (Singh *et al.*, 2017) proposed a technique for handwritten recognition for Kazak and Russian languages. The model used Deep CNNs for feature extraction and Multi Layer Perception (MLP) was used for word classification and the results were compared with a model combining RNN and CNN. The results were compared with Simple HTR recurrent CNN, the recognition accuracy was found to be 75.8%.

The author suggested (Wicaksono & Purwarianti, 2010) a method for offline handwritten text recognition using encoder decoder technique. CNN was used as an encoder for input text line image whereas the Bidirectional LSTM fully connected to CNN was used as a decoder for the sequential prediction of handwritten Greek characters. The newly created EPARCHOS dataset was used and the results were calculated using MSE and MAR flexibility, the final perimeters showed the gradual increase in parameters using above.

2. Material and Methods

Ancient text translators essentially consist of various modules, implemented using deep learning, that convert the source text into a target language. A Neural Network (NN) has been used for the development of the model. The model consists of an encoder and a decoder, which involves running two LSTM Recurrent Neural Networks which work together simultaneously to transform one sequence

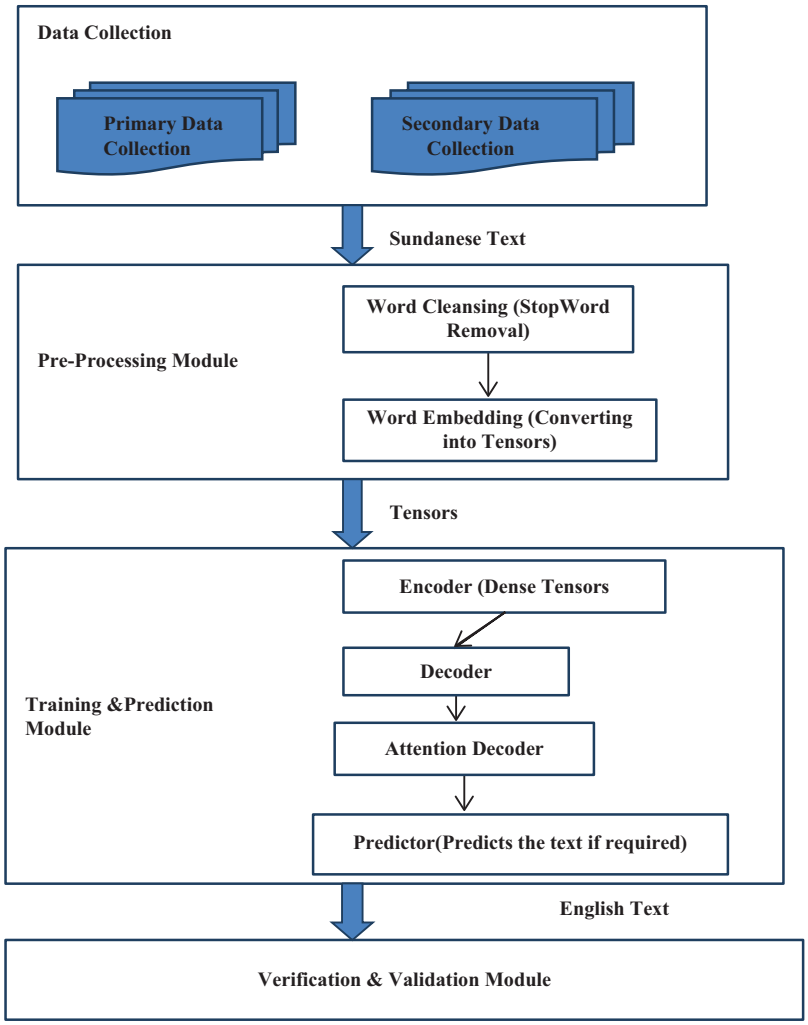

Figure 1. Architecture for Translation

into another. An Encoder network condenses an input sequence into a vector, a Decoder network unfolds that vector into a new sequence.

LSTM is a type of RNN model that computes the probability of occurrence of words in a reference text. The probability of a sequence of T words {w1, ...,wT} is denoted as P(w1, ..., wT). Since the number of words coming before a word, wi , varies depending on its location in the input document, P(w1, ..., wT) is usually conditioned on a window of n previous words rather than all previous words given in equation 1 (Costa et al., 2016). Pytorch is used for deep learning which works by converting the inputs into tensors.

$$P(w1,., wT) = i=T \prod i=1 P(wi \mid w1, .,wi-1) \approx i=T \prod i=1 P(wi \mid wi-(n-1) ,., wi-1) \quad (1)$$

Figure 1 represents the flow for the translation of ancient texts, the first module includes both secondary and primary datasets for data collection. The secondary dataset i.e., Sunda dataset, was obtained from the author (Markou, K. *et al*. (2021), whereas the primary dataset was collected by converting English language into a Sunda dataset using a pre-existing translator. The data was preprocessed using word cleansing and word embedding i.e. converting into tensors.

2.1. Data Collection

In the proposed method of translation, the pre-existing dataset has been collected (Hermanto *et al*., 2015) and a self created corpus has been generated by translating English text into Sundanese using a pre-existing translator, so both secondary and primary data collection was carried out. Due to the unavailability of existing Sundanese to English parallel corpus, the primary data collection was done by generating a parallel corpus, by translating the English text into Sundanese text using Google Translate. The corpus contained 6615 bilingual sentences, 13,230 sentences and 76757 characters. An example of the created parallel corpus is shown in Table 1.

Table 1. Bilingual Corpus of Sundanese to English text.

English Text (Input Text)	i'd rather rather at home go out of milk	i have to make a complaint about	the students from from and and guru	and offered what was left to the little mouse
Sundanese Text (Output Text)	kuringlangkun-gresepcicing di bu-mitinimbangkaluar	Abdikedahnga-damelkeluhanngeu-naan	muriddiajartina guru sarengogé	Sarengmasihanna-onanukéncakabeu-ritsakedik

Furthermore, for secondary data collection, the data collected from (Huang et al., 2018) contains a character level dataset of a size of 6.7 MB, word level annotated dataset of a size of 231 MB with 66 images of the original Sundanese language converted into text format. The used Sundanese text originated from Brahmi Script in 1400 CE in the Java region and consist of 66 classes with 27 consonants, 7 vowels and 10 numerals. An image of the sentence level dataset is shown in Fig.2

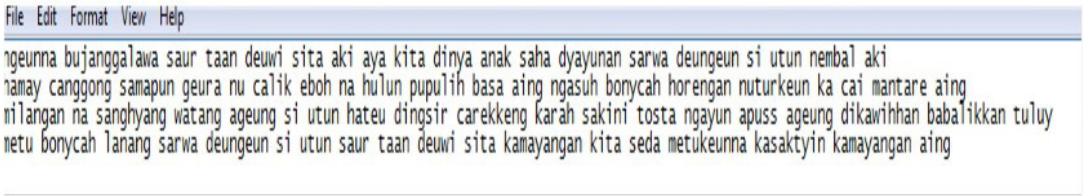

Figure 2. Original Sundanese Text

3. Methodology

The translation and prediction model comprises vectorization and training i.e. taking source text as input and converting it into dense tensors which are further converted into target text with the help of encoder-decoder units. Vectorization further comprises of text cleansing and one-hot encoding, which involves converting the source code i.e. Sundanese text into tensors.

The preprocessing involves text cleansing that reduces the size of the source text by removing stop words, punctuations and finally converting all the upper case into lower case. As the text is very ancient and due to very low level of knowledge of the source language i.e., Sundanese stemming and lemmatization was not be carried out in the proposed technique, thereafter, the size of the text was reduced to 80%, along with the performance of one-hot encoding which involves converting the clean text into tensors.

The training and prediction technique used in this study comprises of three functions, the first function converts characters to numeric indexes so that each character in the text has a unique index, whereas the second function converts those indexes into tensors. Finally, the same thing has been done for the whole pair of English to Sundanese, so the generated output is in the form of one tensor using Pytorch.

After the tensors were created, the training section took the tensors as input and converted them into the target language i.e., English. It was designed using a Neural Network (NN) consisting of an encoder and a decoder, that involved running two RNN simultaneously to transform one sequence to another (Purwarianti *et al.*, 2015). An encoder network condenses an input sequence into a vector, and a decoder network unfolds that vector into a new sequence. The encoder of a seq2seq network is an RNN that outputs some value for every word from the input sentence. For every input word the encoder outputs vectors and a hidden state, and uses the hidden state for the next input word.

Initially, the input is given to the encoder in the form of tensors which is further passed to the embedded layer that is used to give the dense representation to the words and their complicated meaning, input to the layer is integer encoded in order to represent each word uniquely and the same is given to Gated Recurrent Unit (GRU) (Fachrurrozi *et al.*, 2014) which is a kind of LSTM that intends to solve the problem of vanishing gradients, another input to GRU (Markou, K. *et al.*, 2021) is from a previously hidden layer, but as it is the first layer so the previously hidden layer tends to be empty and the result is further transferred to the output layer and hidden layer until the end of the sentence is encountered after which input is given to next layer.

The decoder takes the input from previous hidden layer and output of the encoder which is passed to the embedding layer, furthermore, the embedded output is passed to activation function Rectified Linear Unit activation function (ReLU) to avoid overfitting (Gautam & Chai, 2020) and the output of the function is passed to softmax which calculates a probability for every possible class, thereafter, the output is provided to the hidden layer until the end of the sentence is reached after which the output is provided to the output layer.

The output of the decoder layer is transferred to the attention decoder that is used to improve the performance of the encoder-decoder translation module in parallel, it gets the input from the encoder along with the decoder and a previously hidden layer of the decoder. The input of the decoder layer is given to the embedded layer that again gives a dense representation of tensors, which is further provided to the dropout layer to avoid overfitting, afterwards, this embedded output is given to the attention layer that produces attention weights for each tensor that are further given as input to Batch Matrix Multiplication (BMM) along with embedded output that is further passed to GRU to solve the problem of gradient vanishing and keeps on giving the input to the previously hidden layer until the end of the sentence is encountered. The training and prediction module are used to predict tensors into source language, by using the mapping done by the decoder, to finally convert tensors into a text file of the target language.

The summary of the LSTM based model is given below in Table 2. It consists of Hidden _Size= 256, Input_Size=4345, Output _ Size=3586. The model was trained internally on an 80% dataset and validated on a 20% dataset. The model was tested on 250 sentences and as shown in Fig.3, with a gradual increase of training dataset, the training loss decreases.

Table 2. Summary of the Training Model

Layer(Type)	Output Size	Parameter
conv2d_1 (Conv2D)	(None,32,32,32)	896
conv2d_2 (Conv2D)	(None,30,30,32)	9248
max_pooling2d_1 (MaxPooling 2)	(None,15,15,32)	0
dropout_1 (Dropout)	(None,15,15,32)	0
conv2d_3 (Conv2D)	(None,15,15,64)	18496
conv2d_4 (Conv2D)	(None,13,13,64)	36928
max_pooling2d_2 (MaxPooling 2)	(None,6,6,64)	0
dropout_2 (Dropout)	(None,6,6,64)	0
conv2d_5 (Conv2D)	(None,6,6,64)	36928
conv2d_6 (Conv2D)	(None,4,4,64)	36928
max_pooling2d_3 (MaxPooling 2)	(None,2,2,64)	0
dropout_3 (Dropout)	(None,2,2,64)	0
Flatten_1 (Flatten)	(None, 256)	0

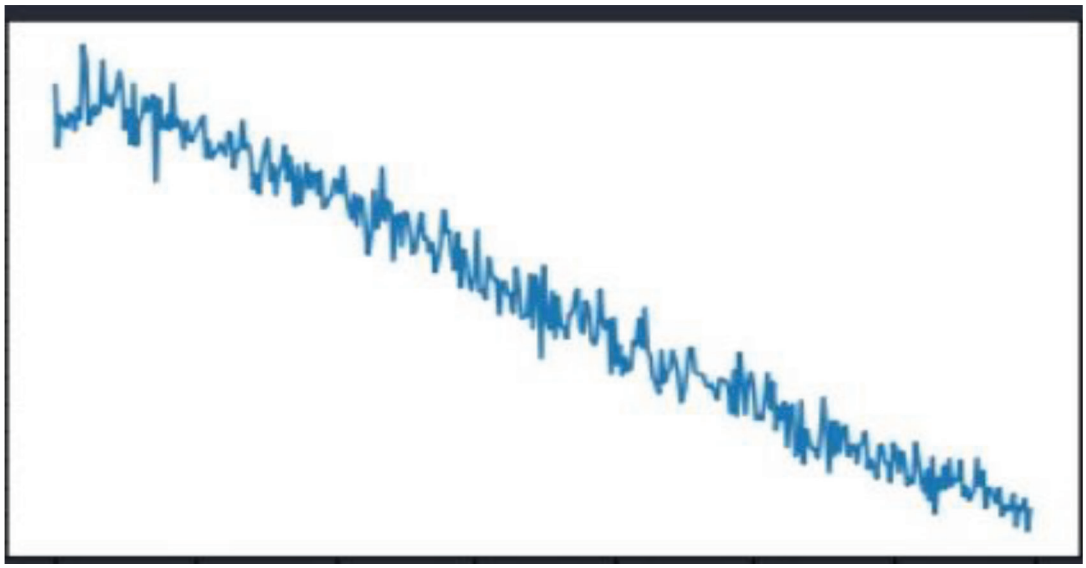

Figure 3. Graph showing Training Loss with 4500 epochs.

4. Results and Discussion

The results of Sundanese to English translation have been evaluated using manual and automatic evaluation. The model proposed by us has been evaluated on 250 sentences and analysis has been performed by evaluating and comparing various parameters, such as Average BLEU score, Word Error Rate WER and N-gram analysis on a pre-existing translator i.e. IM Translator.

4.1. Comparative Analysis

4.1.1. Word Error Rate (WER)

The WER is the difference in length sequence of recognizable text and reference text by Large Vocabulary Continuous Speech Recognition (LVCSR), the word sequence is hypothesized and is aligned with the reference script in order to calculate WER. If there is an N number of total words, the error is calculated as a summation of the number of Substitutions (S), the number of Insertions (I), and number of Deletions (D), WER is calculated by the equation [12].

$$WER = \frac{I + S + D}{N} \times 100 \tag{2}$$

The frequency range has been set by taking the average of Word Error rate for 250 sentences and four ranges, as shown in Fig.4. The WER is calculated, it is then compared with an existing IM Translator shown in Fig 4 and it has been observed that there were more sentences with WER (75-100%) for IM Translator than for Sun Tran, with a difference of approximately 4%. Whereas the total number of sentences with WER (25- 50%) for IM translator was approximately 4% higher than Sun translator, which increases the performance of the Sun Tran.

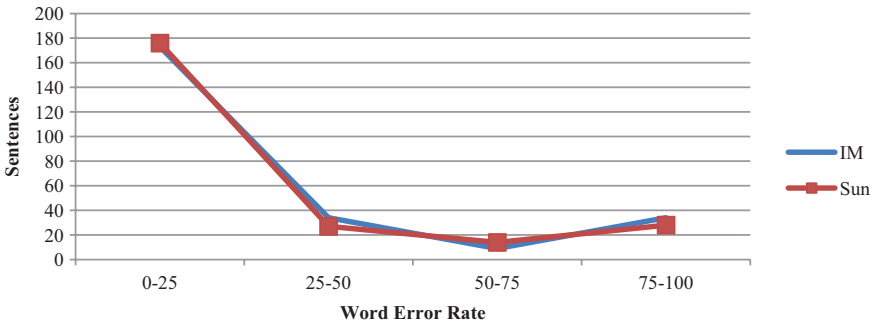

Figure 4. Comparison of Word Error Rate

4.1.2. Average Bleu Score Analysis

BLEU score is an automatic evaluation of the quality of text translated from one language to another, using a reference that was previously translated. The similarity index of the translated value is calculated by counting the same word that appears in both references and translation. The model proposed by us was evaluated on 250 sentences, the BLEU score was calculated, as shown, using equations 3 and 4.

$$B.P = \begin{cases} 1 \; if c > r \\ e\left(1 - \frac{r}{c}\right) if c > r \end{cases} \tag{3}$$

$$BLEU = B.P \times \exp\left(\sum_{n}^{1} W_n \log p_n\right) \tag{4}$$

BP is a Brevity Penalty, which counters the length of translation result considering r as the length of reference text, whereas c is the length of translated text and Pn is precision recall of each n-gram. Average BLEU Score was calculated for Sun Trans and IM Translator in which BLEU score value for Sun Trans was 19 times greater than for IM Trans. Therefore, Sun Tran BLEU score was 8% greater than IM translator, as shown in Fig.5.

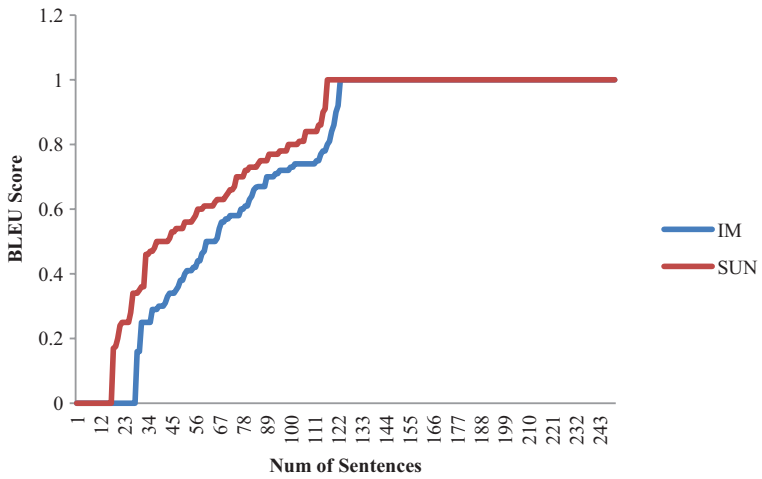

Figure 5. Comparison of Average BLEU Score

4.1.3. N-Gram Analysis

N gram is a continuous sequence of n words from a reference text, which are basically sets of occurring words in the text, when computing, N gram typically moves one word forward. When N=1 it is called 1-gram (uni-gram), if N=2 it is called 2-gram (bi-gram), if N=3 it is called 3-gram (tri-gram), if N is greater than 3 it is called 4-gram or 5-gram analysis is done.

If there are X words in a given sentence K and the number of n-grams in the sentences is computed as per given equation below.

$$N\text{-}gramsK = X\text{-}(N\text{-}1) \tag{5}$$

The Meteor score m is calculated using the mapped unigrams between the two strings i.e., the total number of unigrams in translation and the total number of unigrams in reference are shown in the equation.

$$M = t/r \tag{6}$$

UNI-GRAM

A uni-gram analysis is performed on all the words in a sentence, one by one, such as “kuring”, “langkung” ,“resep” ,“cicing”. In uni-gram analysis it is considered that the occurrence of each word in the sentence is independent of the previous word. The results indicate that for the words with Meteor score 0% been decrease by 3% for SUN-Tran, whereas for sentences with Meteor score 100% increases with 4% than IM –Tran as shown in Fig.6.

BI-GRAM

A bi-gram analysis on sequence of two words, such as “kuringlangkung”, “resepcicing”, was performed. In bi-gram it is assumed that the occurrence of each word is only dependent on the word that preceded it. Hence two words are counted as one gram i.e. one feature. The results indicate that meteor score for the sequence with Meteor score 0% has been decrease by 5% for SUN-Tran whereas for sentences with 100% meteor score increases with 5% than IM –Tran as shown in Fig 6.

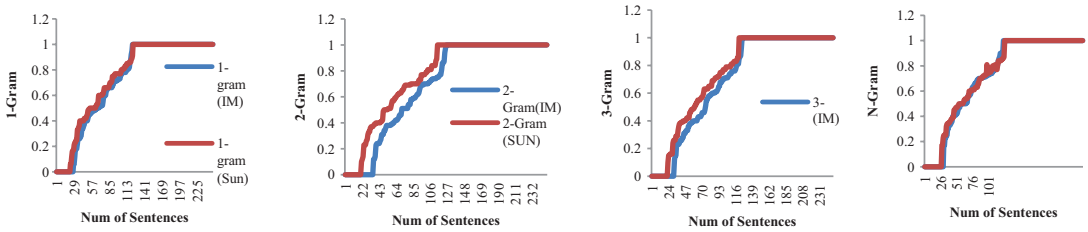

Figure 6. N-gram Analysis for IM and Sun Trans

TRI-GRAM

A tri-gram analysis on a sequence of three words, such as “ayakitadinya” was done. In Tri- Gram it is assumed that the occurrence of the word in the sentences is dependent on its previous two words. In tri-gram three words are taken as 1 gram i.e. one feature. The results has been found that for the words with 0% Meteor score has been decrease by 6% for SUN-Tran whereas for sentences with 100% Meteor score has been increased with 4% than IM –Tran as shown in Fig 6.

N-gram analysis (N>3)

An n-gram analysis is done on a sequence of four or more words, such as “ayakitadinyaanak". In n-gram, a collection of more than 3 words is taken as 1-gram and the occurrence of the word depends upon the N previous words. In an n-gram sequence, N words are taken as one-gram or one feature. The result indicates that for the words with 0% meteor score has been decrease by 2% for SUN-Tran whereas for sentences with 100% Meteor score has been increases with 4% than IM –Tran as shown in Fig.6.

For the construction of the API, as shown in Fig 7, two different layers are mapped using URL path (‘/’) for initial and final webpages. Two HTML template pages are rendered in API. When Python code is run, it connects to the Flask server at ‘http://localhost:5000/’. Flask then checks if there is a similarity match between the provided path and the defined function and shows us our HTML markup.

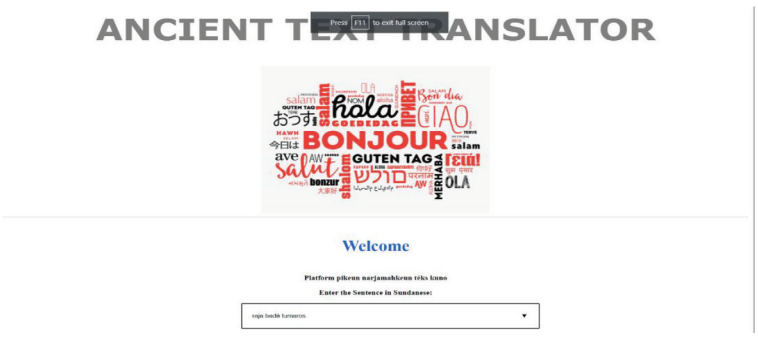

Figure 7. Screenshot of the Sun Translator

First HTML markup caters a home page for welcoming the visitors to the site hosting the API. It contains a dropdown menu to select the desired languages, a text area to enter a sentence and a convert button. Hitting this convert button will make Flask run our Deep Learning Model, redirecting it to our output webpage.

Second HTML markup is a webpage which displays the result returned from our Deep Learning Model. This page ends with a note thanking the user.

5. Conclusions

In comparison to the early translations of Sundanese text into English using the IM translator, it was found that the developed Sun-Tran model achieves more accurate results, as the Average Bleu Score increased by 8%, WER decreased at certain frequency ranges from 14% and 10%. N-gram analysis showed that Uni-gram increased by 4%, Bi-gram increased by 5% and Tri-gram increased by 4% and 4-gram increased by 4% when compared with the previously developed IM Translator for Sundanese to English Translation.

Several problems had to be addressed during the translation of Sundanese text into English due to a limited bilingual corpus and a large number of complex non-translated words, as very little knowledge of the language is available. Furthermore, the translator can also be used to support the translation of low resource source languages to universal languages.

6. Annexure

Abbreviation	Nomenclature
ML	Machine Learning
LSTM	Long Short Term Memory
PBST	Phrase-Based Statistical Translator
RNN	Recurrent Neural Network
LG	Link Grammar
SBTT	Statistical-Based Translation Technique
BLEU	Bilingual Evaluation Under Study
OOV	Out of Word Vocabulary
NMT	Neural Machine Translation
PPR	Precision Pattern Rate
CNN	Convolution Neural Network
GRU	Gated Recurrent Unit
ReLU	Rectified Linear Unit activation function
BMM	Batch Matrix Multiplication
WER	Word Error Rate
HTML	Hyper Text Markup Language
R^2RNN	Recursive Recurrent Neural Network

7. References

Ali and Renals. S, “Word Error Rate Estimation for Speech Recognition: e-WER.” [Online]. Available: https://github.com/qcri/e-wer.

Al-Muzaini. H. A,. Al-Yahya. T. N, & Benhidour. H, “Automatic Arabic Image Captioning using RNN-LSTM-Based Language Model and CNN,” 2018. [Online]. Available: www.ijacsa.thesai.org.

Apriyanti, T., Wulandari, H., Safitri, M., & Dewi, N. (2016). Translating Theory of English into Indonesian and Vice-Versa. In Indonesian Journal of English Language Studies (Vol. 2, Issue 1).

Athiwaratkun and J. W. Stokes, “Malware classification with lstm and gru language models and a character-level CNN.”, 2016.

Bijalwan, V., Kumar, V., Kumari, P., & Pascual, J. (2014). KNN based machine learning approach for text and document mining. International Journal of Database Theory and Application, 7(1), 61–70. https://doi.org/10.14257/ijdta.2014.7.1.06

Chadha, S., Mittal, S., & Singhal, V. (2019). An insight of script text extraction performance using machine learning techniques. International Journal of Innovative Technology and Exploring Engineering, 9(1), 2581–2588. https://doi.org/10.35940/ijitee.A5224.119119.

Chadha, S., S. Mittal, and V. Singhal. "Ancient text character recognition using deep learning." International Journal of Engineering Research and Technology 3.9 (2020): 2177-2184.

Chaudhary. J and Patel. A, “IJSRSET1844500 | Bilingual Machine Translation Using RNN Based Deep Learning,” vol. 4, 2018, [Online]. Available: www.ijsrset.com.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. https://arxiv.org/abs/1406.1078

Fachrurrozi, M., Yusliani, N., & Agustin, M. M. (n.d.). "Identification of Ambiguous Sentence Pattern in Indonesian Using Shift-Reduce Parsing", 2014

G. Lin and W. Shen, “Research on convolutional neural network based on improved Relu piecewise activation function,” in Procedia Computer Science, 2018, vol. 131, pp. 977–984, https://doi.org/10.1016/j.procs.2018.04.239.

Gautam N. & Chai S. (2020). Translation into Pali Language from Brahmi Script. In: Sharma D.K., Balas V.E., Son L.H., Sharma R., Cengiz K. (eds) Micro-Electronics and Telecommunication Engineering. Lecture Notes in Networks and Systems, vol 106. Springer, Singapore. https://doi.org/10.1007/978-981-15-2329-8_12.

Hermanto, A Adji. T, & Setiawan, N (2015)“Recurrent neural network language model for English-Indonesian Machine Translation: Experimental study,” 2015 International Conference on Science in Information Technology (ICSITech), Oct. 2015, doi: https://doi.org/10.1109/icsitech.2015.7407791.

Hinton, G. E., & Zemel, R. S. (n.d.). Autoencoders, Minimum Description Length and Helmholtz Free Energy.

Lauriola, I., Lavelli, A., & Aiolli, F. (2022). An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools. Neurocomputing, 470, 443–456. https://doi.org/10.1016/j.neucom.2021.05.103.

M. R. Costa-Jussà and J. A. R. Fonollosa, “Character-based Neural Machine Translation,” Mar. 2016, [Online]. Available: https://arxiv.org/abs/1603.00810.

M. Suryani, E. Paulus, S. Hadi, U. A. Darsa, and J. C. Burie, "The Handwritten Sundanese Palm Leaf Manuscript Dataset from 15th Century," in Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, Jul. 2017, vol. 1, pp. 796–800, doi: 10.1109/ICDAR.2017.135.

M. Zhang, Y. Zhang, and D.-T. Vo, "Gated Neural Networks for Targeted Sentiment Analysis." [Online]. Available: www.aaai.org.

Maitra, D. sen, Bhattacharya, U., & Parui, S. K. (2015). CNN based common approach to handwritten character recognition of multiple scripts. Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, 2015-Novem, 1021–1025. https://doi.org/10.1109/ICDAR.2015.7333916

Markou, K. et al. (2021). A Convolutional Recurrent Neural Network for the Handwritten Text Recognition of Historical Greek Manuscripts. In: , et al. Pattern Recognition. ICPR International Workshops and Challenges. ICPR 2021. Lecture Notes in Computer Science(), vol 12667. Springer, Cham. https://doi.org/10.1007/978-3-030-68787-8_18.

Miyamoto, Y., & Cho, K. (2016). Gated Word-Character Recurrent Language Model. https://arxiv.org/abs/1606.01700.

Nurseitov.D, Bostanbekov.K, Kanatov.M, Alimova.A 1,2 , Abdallah.A, Abdimanap.G (2021). "Classification of handwritten names of cities and Handwritten text recognition using various deep learning models", Advances in Science, Technology and Engineering Systems Journal Vol. 5.

P. Wang, P. Nakov and H. T. Ng, "Source Language Adaptation Approaches for Resource-Poor Machine Translation," 2016, doi: 10.1162/COLI.

P. Zhou, Z. Qi, S. Zheng, J. Xu, H. Bao, and B. Xu, "Text Classification Improved by Integrating Bidirectional LSTM with Two-dimensional Max Pooling," Nov. 2016, [Online]. Available: https://arxiv.org/abs/1611.06639.

P.Y. Huang, F. Liu, S.-R. Shiang, J. Oh, and C. Dyer, "Attention-based Multimodal Neural Machine Translation.", 2018.

Purwarianti.A, Yayat.D, Fakultas.P, "Experiment on a Phrase-Based Statistical Machine Translation Using PoS Tag Information for Sundanese into Indonesian", International Conference on Information Technology Systems and Innovation (ICITSI),p.p 1-6, 2015.

R. P. Haroon and T. A. Shaharban, "Malayalam machine translation using hybrid approach," 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), Mar. 2016, doi: 10.1109/iceeot.2016.7754839.

Ray, A., Rajeswar, S., & Chaudhury, S. (2015). Text recognition using deep BLSTM networks. ICAPR 2015 - 2015 8th International Conference on Advances in Pattern Recognition. https://doi.org/10.1109/ICAPR.2015.7050699

Ray, A., Rajeswar, S., & Chaudhury, S. (2015). Text recognition using deep BLSTM networks. ICAPR 2015 - 2015 8th International Conference on Advances in Pattern Recognition. https://doi.org/10.1109/ICAPR.2015.7050699

S. P. Singh, A. Kumar, H. Darbari, L. Singh, A. Rastogi and S. Jain, "Machine translation using deep learning: An overview," 2017 International Conference on Computer, Communications and Electronics (Comptelix), 2017, pp. 162-167, https://doi.org/10.1109/COMPTELIX.2017.8003957. Results should be clear and concise.

Singh. S, Kumar. A, Darbari. H, Singh.L, Rastogi. A & Jain.S, (2017). "Machine translation using deep learning: An overview," 2017 International Conference on Computer, Communications and Electronics (Comptelix), Jul. 2017, https://doi.org/10.1109/comptelix.2017.8003957.

Swe, T & Tin. P. (2005). Recognition and Translation of the Myanmar Printed Text Based on Hopfield Neural Network. 6th Asia-Pacific Symposium on Information and Telecommunication Technologies, APSITT 2005 - Proceedings. 2005. 99 - 104. https://doi.org/10.1109/APSITT.2005.203638.

Swe, T., & Tin, P. (n.d.). Recognition and Translation of the Myanmar Printed Text Based on Hopfield Neural Network.

Wicaksono, A and Purwarianti. A, "HMM Based Part-of-Speech Tagger for Bahasa Indonesia Implementing Deep Learning Using Sequence-to-sequence for Automatic Question Generator View project Game refinement theory (M/P) View project HMM Based Part-of-Speech Tagger for Bahasa Indonesia," 2010. [Online]. Available: https://students.itb.ac.id/home/alfan_fw@students.itb.ac.id/IPOSTAgger.

Windu, M., Kesiman, A., Burie, J.-C., & Ogier, J.-M. (2016). A New Scheme for Text Line and Character Segmentation from Gray Scale Images of Palm Leaf Manuscript. https://doi.org/10.1109/ICFHR.2016.63

Xiaoyuan.Y, Ruoyu. L and Maosong. S, (2017)."Generating Chinese Classical Poems with RNN Encoder-Decoder", China National Conference on Chinese Computational Linguistics, pp 211-222, 2017.