IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 10, October 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023. 1210063 |

Next Word Predictor in Gujarati Language

Shail Patel, Karan Patwa, Deepak C. Vegda, Sunil K. Vithlani

U.G. Student, Department of Information Technology, Dharmsinh Desai University, Nadiad, Gujarat, India

U.G. Student, Department of Information Technology, Dharmsinh Desai University, Nadiad, Gujarat, India

Asst. Professor, Department of Information Technology, Dharmsinh Desai University, Nadiad, Gujarat, India

Asst. Professor, Department of Information Technology, Dharmsinh Desai University, Nadiad, Gujarat, India

ABSTRACT: There are many local languages spoken in India such as Hindi, Marathi, Kannad, Bengali, Gujarati etc. These languages are little known throughout the world as well as they are complex to understand and use. As the languages have complex grammar rules, it is difficult to even predict the next word let alone generating a whole sentence and hence there is minimal amount of work done in this area. Next word prediction can be called a primitive part of a much larger task of sentence generation. This paper presents Recurrent Neural Networks (RNNs), which are a type of neural network that can learn to predict the next word in a sequence. RNNs work by saving the output of each layer and feeding it back into the input of the next layer. This allows them to learn the long-term dependencies between words, which is essential for understanding natural language. In this research, the RNN is used to predict the next word in Gujarati language. It works with an untagged plain text dataset. We extrapolated that with increase in size of dataset the model can give outputs become more and more contextualized. However, as the model works in unidirectional way, it cannot produce the word based on context present in other direction. So, bidirectional models can be used to improve the accuracy to a certain extent.

KEYWORDS: Natural Language Processing (NLP), Text Generation, Recurrent Neural Network (RNN), Gujarati Text, Next Word Prediction.

I. INTRODUCTION

NLP is a sub-domain of Machine Learning which deals with eclectic range of applications. Natural language processing (NLP) is the ability of a computer program to understand human language as it is spoken and written referred to as natural language. It is a component of artificial intelligence (AI). The NLP has its roots in field of linguistics since 1940s. According to paper [14], NLP has it's history in 4 phases: phase 1 (Late 1940s to Late 1960s), phase 2 (Late 1960s to 1970s), Phase 3 (Late 1970s to Late 1980s) and phase 4 (Late 1980s to 2000s) and currently it can be said that the 5th phase is ongoing. It has a variety of real-world applications in a number of fields, including but not limited to medical research, search engines and business intelligence.

Due to NLP, computers can understand natural language as humans do. Whether the language is spoken or written, natural language processing uses artificial intelligence which enables computer to take real-world input, process it and make sense out of it. Humans have different sensory organs such as ears to hear and eyes to see similarly, computers have programs to read and microphones to collect audio. And just as humans have a brain to process that input, computers have a program to process their respective inputs. At some point in processing, the input is converted to code that the computer can understand [4].

There are two main phases to natural language processing: data pre-processing and algorithm development [2]. Data pre-processing involves preparing and "cleaning" text data for machines to be able to analyse it. Pre-processing puts data in workable form and highlights features in the text that an algorithm can work with.

Once the data has been pre-processed, an algorithm is commenced to process it. There are several different natural language processing algorithms, however two significant types are commonly used:

- Rules-based system [8]: This system uses meticulously designed linguistic rules. This approach was used early on in the development of natural language processing, and is still used.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023. 1210063 |

- Machine learning-based system [9]: Machine learning algorithms use statistical methods. They learn to perform tasks based on training data they are fed, and adjust their methods as more data is processed. Using a combination of machine learning, deep learning and neural networks, natural language processing algorithms hone their own rules through repeated processing and learning.

There are various stages of natural language processing: Tokenization, Lexical Analysis, Syntactic Analysis, Semantic Analysis, Pragmatic Analysis and Knowledge Discovery. In tokenization stage, text segmentation within the document is performed. Lexical analysis focuses on discovering the data that might need additional processing. Semantic analysis extracts the word sequences and determines syntactic or grammatical structure in each sentence. Semantic analysis is used for providing a structured frame of words. Pragmatic analysis analyses the coherency of words based on their meaning. In knowledge discovery stage, the valuable information from a sophisticated and unstructured text is discovered [1].

Problem Definition

The purpose of this research is to predict the next word using RNN model for Gujarati language, based on the outputs of this research work, the enhancement of RNN model or usage of another model for the purpose of sentence generation can be proposed.

Aims and Objectives

The main objective of this research is to measure the extent to which the RNN model is able to predict the next word in Gujarati language. This research work also focuses on comparing the results of the various types and sizes of dataset used to train the model, provide the possible problems faced commonly by the model and give the prospects of future enhancements available.

Problems and Challenges

There is little to no work done when it comes to working with NLP in Indic Languages [11] especially in Gujarati language and so it became difficult to handle various exceptions and errors encountered during the implementation. The most notable challenge we encountered was locating and accessing Gujarati text dataset and cleaning it. Also, there is no work done in Part-of-speech(POS) Tagging [13] the Gujarati words which makes it even harder to implement supervised learning algorithms.

Significance and Advantages

This research can be used as a stepping stone to solving a larger goal of sentence or text generation for Gujarati language. This can further be used for writing essays, letters or messages. If the unidirectional RNN is able to generate somewhat contextualized word, the bidirectional RNN can be used to predict the next word not only based on the past context but also considering the context of the text appearing later in the sentence [7].

Paper is organised as follows. Section II describes a couple of problems with early text generation models and the use of RNN along its workflow diagram with explanation. Section III explains the 6 major steps followed for implementing the solution along with sample dataset and training summary of the model. The graphs and images in Section IV shows the results of inputs tested. Lastly, Section V concludes the paper and also provides the future scopes of improvement.

II. LITERATURE SURVEY

Text generation is a sub-field of natural language processing. It leverages knowledge in computational linguistics and artificial intelligence to automatically generate natural language texts, which can satisfy certain communicative requirements. The task of predicting next word from any given set of words is a sub part of natural language text generation. It can be called the most primitive form of text generation. There are various text generation models available. Neural, encoder-decoder models have had significant success in text generation. But, there are various problems which remains unaddressed with this type of generation. There are two major problems (a) uninterpretable

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023. 1210063 |

and (b) difficult to control in terms of their phrasing and content [6]. This research work proposes use of RNN model for predicting the next word.

A recurrent neural network (RNN) represents a class of artificial neural networks designed to handle sequential or time-series data. RNNs are a key component of deep learning and find wide applicability in tasks involving temporal relationships, such as language translation, natural language processing (NLP), and speech recognition. Unlike traditional deep neural networks, RNNs introduce a concept of "memory." They incorporate prior outputs as part of the input to the subsequent layer, enabling the model to generate future outputs based on historical context. This fundamental distinction sets RNNs apart from their non-recurrent counterparts, where inputs and outputs are treated as independent entities. However, it's important to note that unidirectional RNNs have limitations when it comes to considering future events in their predictions [3]. The schematic depiction of an RNN's operation is illustrated in Fig. 1, with further details provided below [5].

- Initial input is image acquired from image pre-processing stage and its feature extraction value has already acquired. – State vector input. State Vector is the desired target from the training image and testing image. State vector can be calculated using formula: $x(k) \epsilon \kappa n$
- Calculate the weight of initial value using the equation: $W1.k,W2.k \epsilon Rmxn$
- Calculate the value of hidden layer using formula: $V1.k,V2.k \epsilon Rmxn$
- Calculate value of ϕ and σ using sigmoid function below:
  $\phi(g(\mu)) = 0.2/(1 + e(-0.2)g(\mu)) - 0.05$
  $\sigma(g(\mu)) = 2/(1 + e(-0.2)g(\mu)) - 0.05$
- Calculate the value of network output using the equation:
  $\mu(k + 1) = A\mu(k) + V\ 1.k\sigma[w, k\mu(k)] + V\ 2.k\phi[W, 2\mu(k)]$

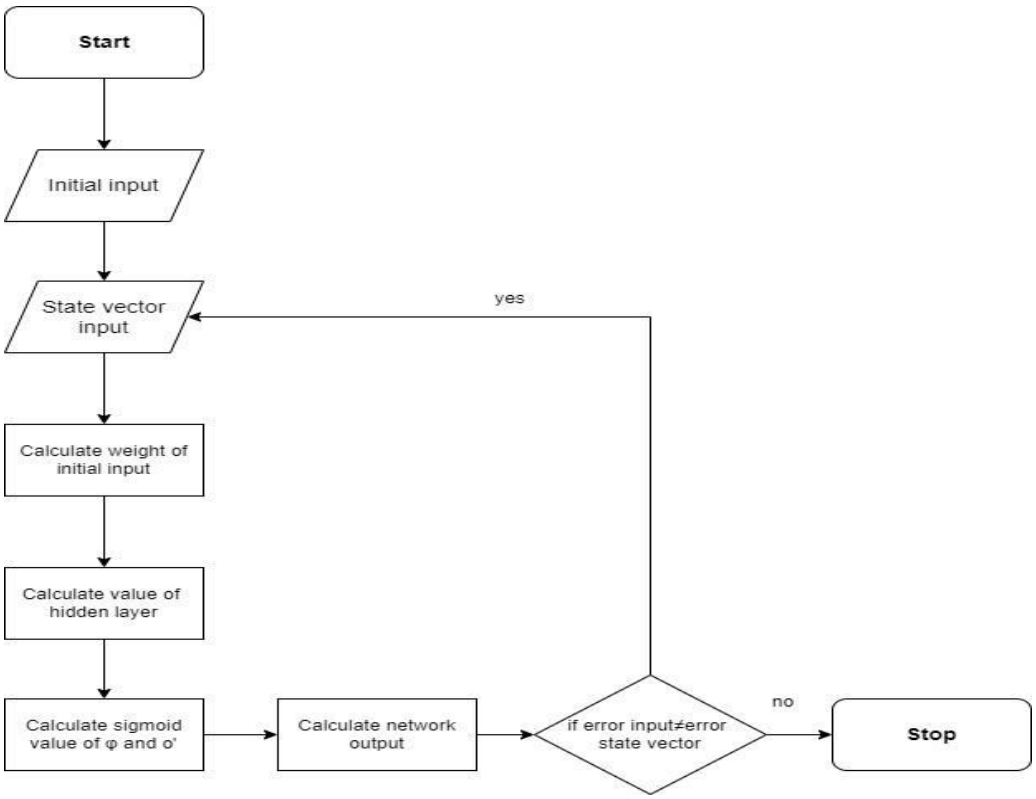


Fig. 1: RNN Flow Diagram [5]

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023. 1210063 |

During the whole research process inspiration and ideas were taken from following research work based on natural language processing and text generation:

The "A Survey of usages of deep learning for Natural Language Processing" [15] shows a detailed explanation of various uses of deep learning inn NLP. This helped us decide to work on text generation. But, there is a lot of work done in that field for English language. So, finally we decided to go for text generation in Gujarati.

For text generation, many state-of-the-arts models are available when it comes to English language. The "BLUERT: Learning Robust Metrics for Text Generation" [16] and "Story Scrambler - Automatic Text Generation Using Word Level RNN-LSTM" [17] discusses about 2 such models: BLUERT and RNN. BLUERT is extended version of BERT high accuracy and is complex to implement. RNN is a comparatively simple and easier to implement. RNN also had a significant accuracy. Also, the time taken for custom training BLUERT is comparatively higher with respect to RNN. So, we used RNN to implement most primitive type of text generation i.e. next word prediction.

III. Proposed Methodology and Discussion

The various steps followed for the implementation of this research work is shown in the fig. 2.

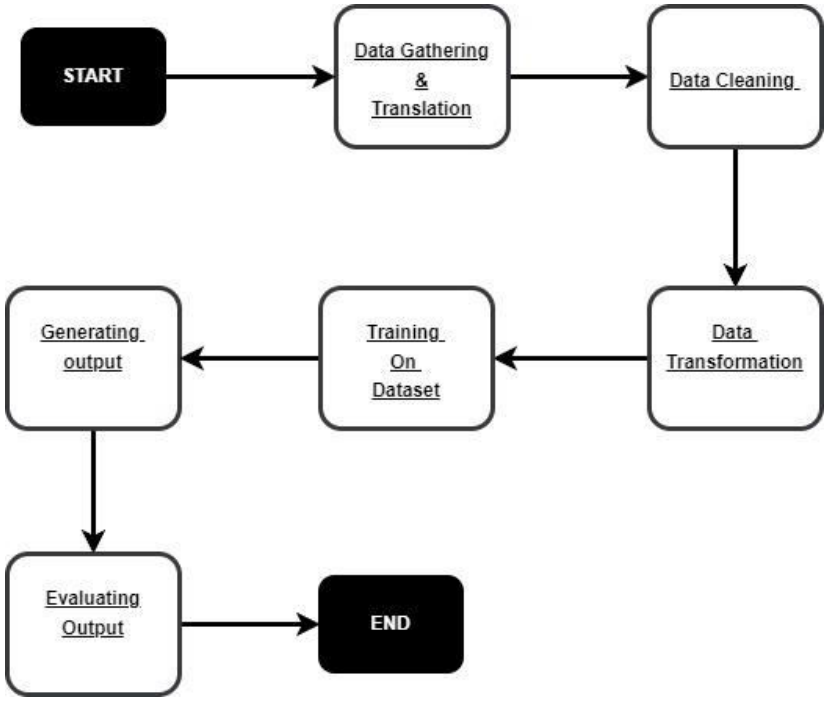


Fig. 2: Implementation Steps

Implementation Details

The model has been prepared and designed on jupyter notebook on a machine with 16 GB LPDD4 RAM, 512 GB SSD and Nvidia GeForce RTX 3050Ti VRAM 4GB GPU.

Dataset Creation

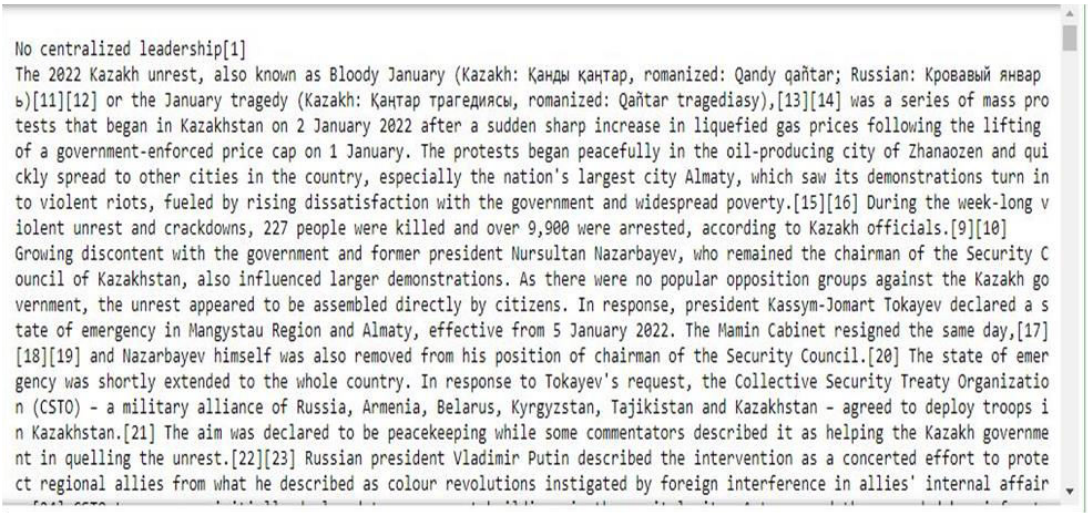
The most significant problem we faced for the implementation of this research work was lack of dataset for Gujarati language. The dataset used here is brought together by 2 techniques. First is that a dataset of any length is generated by fetching the Wikipedia articles of a few topics in English language and translating it to Gujarati. To increase the

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

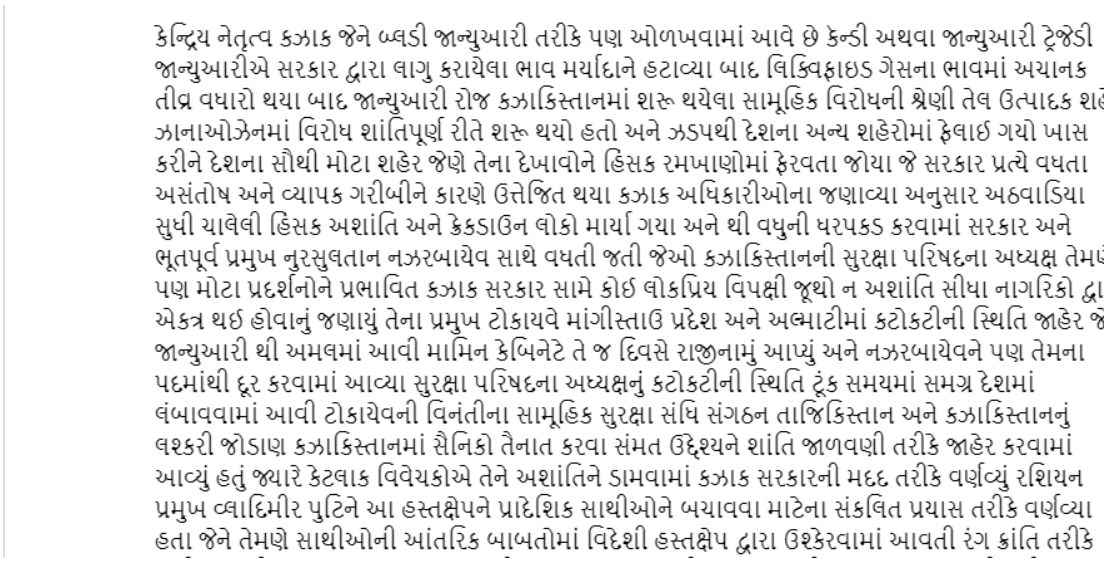
| DOI:10.15680/IJIRSET.2023. 1210063 |

amount of data fetched from a topic, we came up with a code snippet that can help and fetch the titles of various links in the article of topic selected. Each of these titles are used to fetch an article, which is related to the main topic selected and facilitate the increase in the size of the dataset. It is easy to implement once you have the code but the results can have some inaccuracies. Another way used is explicitly accessing the text from various story and prose books written in Gujarati language. This method gives more accurate data but requires a lot of manual work and is time consuming. The total size of dataset is around 16MB consisting of approximately 63000 lines of Gujarati text data, which is after cleaning and removing the noise from the data. Fig. 3,4 shows a part of dataset used in the implementation of the models used in our research work.



No centralized leadership[1]
The 2022 Kazakh unrest, also known as Bloody January (Kazakh: Қанды қаңтар, romanized: Qandy qañtar; Russian: Кровавый январь)[11][12] or the January tragedy (Kazakh: Қаңтар траегдиясы, romanized: Qañtar tragediasy),[13][14] was a series of mass protests that began in Kazakhstan on 2 January 2022 after a sudden sharp increase in liquefied gas prices following the lifting of a government-enforced price cap on 1 January. The protests began peacefully in the oil-producing city of Zhanaozen and quickly spread to other cities in the country, especially the nation's largest city Almaty, which saw its demonstrations turn into violent riots, fueled by rising dissatisfaction with the government and widespread poverty.[15][16] During the week-long violent unrest and crackdowns, 227 people were killed and over 9,900 were arrested, according to Kazakh officials.[9][10]
Growing discontent with the government and former president Nursultan Nazarbayev, who remained the chairman of the Security Council of Kazakhstan, also influenced larger demonstrations. As there were no popular opposition groups against the Kazakh government, the unrest appeared to be assembled directly by citizens. In response, president Kassym-Jomart Tokayev declared a state of emergency in Mangystau Region and Almaty, effective from 5 January 2022. The Mamin Cabinet resigned the same day,[17][18][19] and Nazarbayev himself was also removed from his position of chairman of the Security Council.[20] The state of emergency was shortly extended to the whole country. In response to Tokayev's request, the Collective Security Treaty Organization (CSTO) - a military alliance of Russia, Armenia, Belarus, Kyrgyzstan, Tajikistan and Kazakhstan - agreed to deploy troops in Kazakhstan.[21] The aim was declared to be peacekeeping while some commentators described it as helping the Kazakh government in quelling the unrest.[22][23] Russian president Vladimir Putin described the intervention as a concerted effort to protect regional allies from what he described as colour revolutions instigated by foreign interference in allies' internal affair

Fig. 3: English Dataset



કેન્દ્રિય નેતૃત્વ કઝાક જેને બ્લડી જાન્યુઆરી તરીકે પણ ઓળખવામાં આવે છે કેન્ડી અથવા જાન્યુઆરી ટ્રેજેડી જાન્યુઆરીએ સરકાર દ્વારા લાગુ કરાયેલા ભાવ મર્યાદાને હટાવ્યા બાદ લિક્વિફાઇડ ગેસના ભાવમાં અચાનક તીવ્ર વધારો થયા બાદ જાન્યુઆરી રોજ કઝાકિસ્તાનમાં શરૂ થયેલા સામૂહિક વિરોધની શ્રેણી તેલ ઉત્પાદક શહેર ઝાનાઓઝેનમાં વિરોધ શાંતિપૂર્ણ રીતે શરૂ થયો હતો અને ઝડપથી દેશના અન્ય શહેરોમાં ફેલાઇ ગયો ખાસ કરીને દેશના સૌથી મોટા શહેર જેણે તેના દેખાવોને હિસક રમખાણોમાં ફેરવતા જોયા જે સરકાર પ્રત્યે વધતા અસંતોષ અને વ્યાપક ગરીબીને કારણે ઉત્તેજિત થયા કઝાક અધિકારીઓના જણાવ્યા અનુસાર અઠવાડિયા સુધી ચાલેલી હિસક અશાંતિ અને ક્રેકડાઉન લોકો માર્યા ગયા અને થી વધુની ધરપકડ કરવામાં સરકાર અને ભૂતપૂર્વ પ્રમુખ નુરસુલતાન નઝરબાયેવ સાથે વધતી જતી જેઓ કઝાકિસ્તાનની સુરક્ષા પરિષદના અધ્યક્ષ તેમણે પણ મોટા પ્રદર્શનોને પ્રભાવિત કઝાક સરકાર સામે કોઇ લોકપ્રિય વિપક્ષી જૂથો ન અશાંતિ સીધા નાગરિકો દ્વારા એકત્ર થઇ હોવાનું જણાયું તેના પ્રમુખ ટોકાયવે માંગીસ્તાઉ પ્રદેશ અને અલ્માટીમાં કટોકટીની સ્થિતિ જાહેર જે જાન્યુઆરી થી અમલમાં આવી મામિન કેબિનેટે તે જ દિવસે રાજીનામું આપ્યું અને નઝરબાયેવને પણ તેમના પદમાંથી દૂર કરવામાં આવ્યા સુરક્ષા પરિષદના અધ્યક્ષનું કટોકટીની સ્થિતિ ટૂંક સમયમાં સમગ્ર દેશમાં લંબાવવામાં આવી ટોકાયેવની વિનંતીના સામૂહિક સુરક્ષા સંધિ સંગઠન તાજિકિસ્તાન અને કઝાકિસ્તાનનું લશ્કરી જોડાણ કઝાકિસ્તાનમાં સૈનિકો તૈનાત કરવા સંમત ઉદ્દેશ્યને શાંતિ જાળવણી તરીકે જાહેર કરવામાં આવ્યું હતું જ્યારે કેટલાક વિવેચકોએ તેને અશાંતિને ડામવામાં કઝાક સરકારની મદદ તરીકે વર્ણવ્યું રશિયન પ્રમુખ વ્લાદિમીર પુટિને આ હસ્તક્ષેપને પ્રાદેશિક સાથીઓને બચાવવા માટેના સંકલિત પ્રયાસ તરીકે વર્ણવ્યા હતા જેને તેમણે સાથીઓની આંતરિક બાબતોમાં વિદેશી હસ્તક્ષેપ દ્વારા ઉશ્કેરવામાં આવતી રંગ ક્રાંતિ તરીકે

Fig. 4: Converted Dataset

Dataset Cleaning

As the data gathered contained some inaccuracies, it needed to be cleaned which was again a problem whose solution was not readily available. It required the use of regular expressions implemented using RegEx module of python to remove all the noisy and inaccurately translated data. After this, the pre-processed dataset ready to be transformed and used for training is obtained.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023. 1210063 |

Dataset Transformation

The dataset obtained in text form needs to be converted to some numerical form before it is given to model for training. This task can be done using various methods. The method used depends on the type of model one is using. RNN supports two methods: one-hot encoding [12] and word embedding. We have used word embedding technique which represents each word as a vector of real numbers. The size of vector is typically smaller than the number of unique words.

Training Model on Dataset

The transformed dataset is now ready to be trained. Before training, the model is defined and compiled. The RNN model is trained on 3 different sized datasets. Training time required for datasets of size 51KB, 2.3MB and 3.4MB is 29 seconds, 7 minutes and 8 minutes 1 second respectively. The summary of training for all the three datasets is as shown in Fig 5-7.

Dataset:- data5.txt size:-51 KB
Length of text: 20702 characters
Model: "my_model"

Layer (type)	Output Shape	Param #
embedding (Embedding)	multiple	17408
gru (GRU) s	multiple	3938304
dense (Dense)	multiple	69700

Total params: 4,025,412
Trainable params: 4,025,412
Non-trainable params: 0

Fig. 5: Summary of smallest dataset

Dataset:- data4.txt size:-2.3MB
Length of text: 945384 characters
Model: "my_model_1"

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	multiple	89344
gru_1 (GRU)	multiple	3938304
dense_1 (Dense)	multiple	357725

Total params: 4,385,373
Trainable params: 4,385,373
Non-trainable params: 0

Fig. 6: Summary of mid-sized dataset

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023. 1210063 |

Dataset:- data2.txt size:-3.4MB

Length of text: 1327993 characters

Model: "my_model_2"

Layer (type)	Output Shape	Param #
embedding_2 (Embedding)	multiple	16896
gru_2 (GRU)	multiple	3938304
dense_2 (Dense)	multiple	67650

Total params: 4,022,850
Trainable params: 4,022,850
Non-trainable params: 0

Fig. 7: Summary of largest dataset

Predicting Next Word

After the training of models is completed, the seed input words are provided to the predict method of the trained model. Based on these words, the next word prediction is carried out.

Evaluating Predicted Words

Once the model is able to generate the output, it is provided with various different inputs and is made to generate the output predictions. These predicted words are then observed to check how accurately the model is able to predict somewhat relevant words.

IV. EXPERIMENTAL RESULTS

The loss vs epoch graphs and accuracy vs epoch graphs are shown in Fig 8-10 and the sentences generated are shown in Fig 11-13.

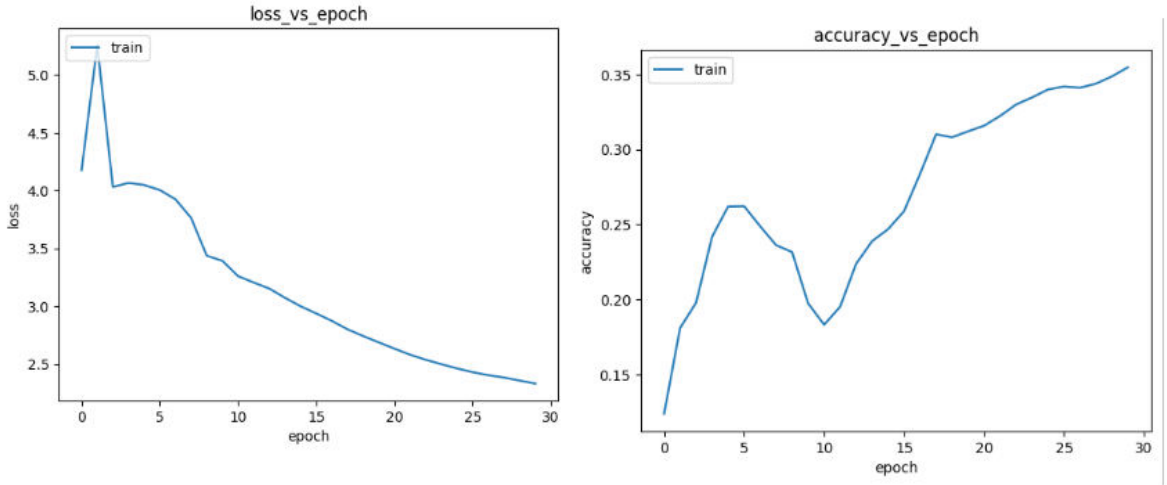


Fig. 8: loss and accuracy vs epoch for smallest dataset

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023. 1210063 |

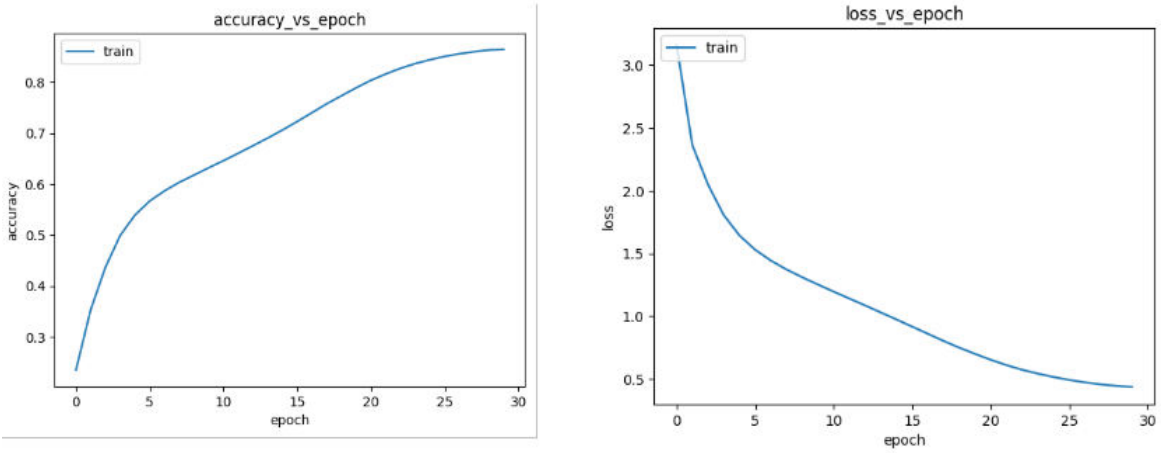


Fig. 9: loss and accuracy vs epoch for mid-sized dataset

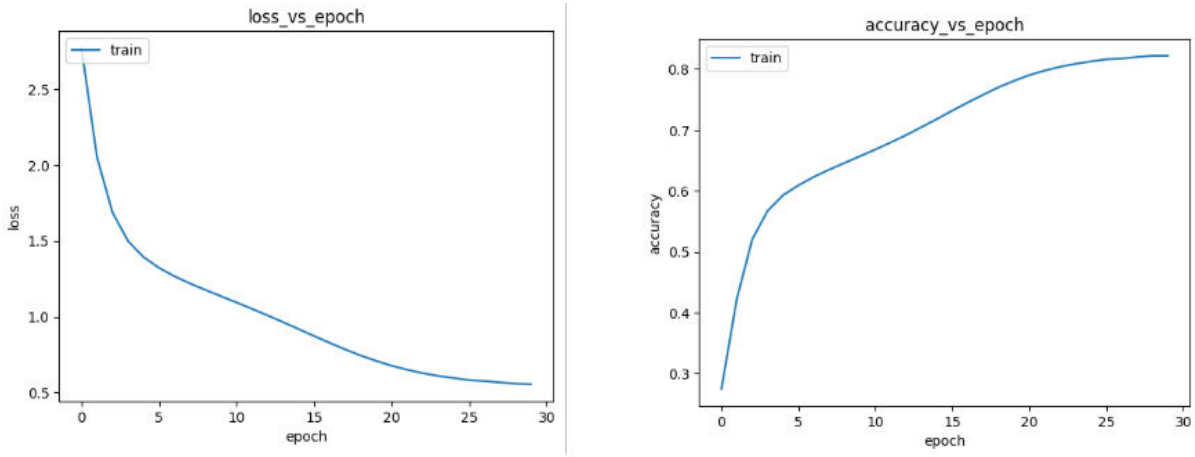


Fig. 10: loss and accuracy vs epoch for largest dataset

Discussion Of Results

The output obtained from RNN observed varying results based on the type and size of dataset used. The predicted word given by model trained on smallest dataset, as seen in Fig. 11, does not make much sense whereas, the outputs given by model for other two datasets as seen in Fig. 12-13 does make some sense. Thus with smaller dataset, the model is not able to provide a proper word let alone the word being contextualized. So, as the size of the dataset increases, the results become more and more accurate. Similar trend can also be seen with the increasing variety of dataset used. By variety we mean that there is myriad category of topics in dataset. It was also observed that if the length of input text is longer, RNN is not able to provide a contextualized result which may be due to unidirectional nature of RNN.

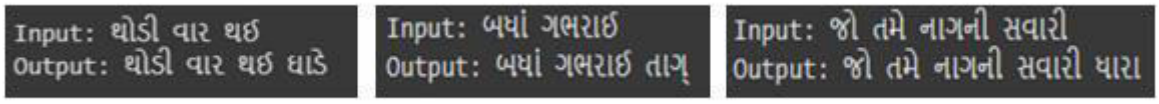


Fig. 11: RNN output for smallest dataset

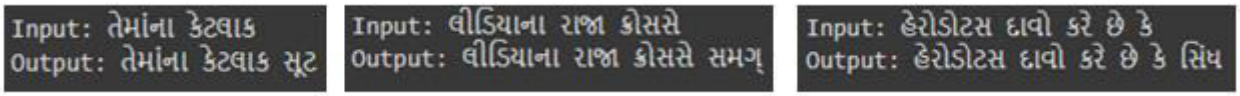


Fig. 12: RNN output for middle-sized dataset

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023. 1210063 |

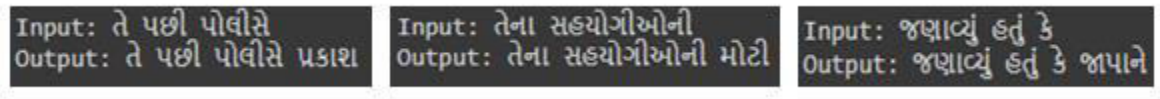


Fig. 13: RNN output for largest dataset

V. CONCLUSION

The results obtained from our experiment are quite acceptable for a unidirectional model. But, the lack of proper datasets in Gujarati language makes it difficult to implement the models that perform really well on large amounts of data. This also hinders the use of supervised models which may provide much more accurate results. Solving the problem of availability of POS Tagged datasets will allow us to explore the approach of supervised learning which has high chances of providing better results.

Despite these impediments it can be said that with proper resources and datasets available, various other models can be implemented and tested. These models can then be modified as per the syntax and grammar of Gujarati language.

Future Scope

There is a scope of improving the dataset and the models can be trained on the improved dataset. This research has been carried out on a very small dataset as there is paucity of Gujarati dataset and limitations of the hardware and resources available, as the hardware available to us was not able to carry out computations for larger dataset. Also, the model currently used is able to predict only a single word or state. This model can be modified to generate text of longer length. Also, as this model will be modified, it can be used to generate essays or stories or letters which are human-like. The other deep learning models like Long-short Term Memory(LSTMs), Convolutional Neural Networks(CNNs), Variational Auto-Encoders(VAEs) and Generative Adversarial Networks(GANs) which perform well for languages like English [10] can be used to perform text generation in Gujarati and compare the results.

REFERENCES

[1] Mahmoudzadeh, Ahmadreza, et al. “Designing Transit Agency Job Descriptions for Optimal Roles: An Analytical Text-Mining Approach.” International Conference on Transportation and Development 2020. Reston, VA: American Society of Civil Engineers.
[2] Lutkevich, Ben. “Natural Language Processing.”, https: // www.techtarget.com / searchenterpriseai / definition / natural-language-processing-NLP.
[3] https://www.ibm.com/topics/recurrent-neural-networks.
[4] Meurers, Detmar. “Natural language processing and language learning.” Encyclopedia of applied linguistics.
[5] Lubis, I., et al. “Image processing method in implementation of handwriting identification for Japanese katakana characters.” Journal of Physics: Conference Series. Vol. 1116. No. 2. IOP Publishing.
[6] Wiseman, Sam, Stuart M. Shieber, and Alexander M. Rush. “Learning neural templates for text generation.” arXiv preprint arXiv: 1808.10122.
[7] Berglund, Mathias, et al. “Bidirectional recurrent neural networks as generative models.” Advances in neural information processing systems 28 (2015).
[8] Hayes-Roth, Frederick. “Rule-based systems.” Communications of the ACM 28.9.
[9] Khan, Wahab, et al. “A survey on the state-of-the-art machine learning models in the context of NLP.” Kuwait journal of Science 43.4.
[10] Iqbal, Touseef, and Shaima Qureshi. “The survey: Text generation models in deep learning.” Journal of King Saud University-Computer and Information Sciences 34.6.
[11] Harish, B. S., and R. Kasturi Rangan. “A comprehensive survey on Indian regional language processing.” SN Applied Sciences 2.7.
[12] Wang, Dongsheng. “Semantic representation and inference for nlp.” arXiv preprint arXiv:2106.08117.
[13] Martinez, Angel R. “Part-of-speech tagging.” Wiley Interdisciplinary Reviews: Computational Statistics 4.1.
[14] Jones, Karen Sparck. “Natural language processing: a historical review.” Current issues in computational linguistics: in honour of Don Walker.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023. 1210063 |

[15] Otter, Daniel W., Julian R. Medina, and Jugal K. Kalita. "A survey of the usages of deep learning for natural language processing." IEEE transactions on neural networks and learning systems 32.2.
[16] Sellam, Thibault, Dipanjan Das, and Ankur P. Parikh. "BLEURT: Learning robust metrics for text generation." arXiv preprint arXiv.
[17] Pawade, Dipti, et al. "Story scrambler-automatic text generation using word level RNN-LSTM." International Journal of Information Technology and Computer Science (IJITCS) 10.6

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.423

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details