International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6621
Article Received: 01 March 2023 Revised: 15 April 2023 Accepted: 01 May 2023

Named Entity Recognition for English Language Using Deep Learning Based Bi Directional LSTM-RNN

Sanjay Kumar Duppati[1], Dr. A.Ramesh Babu[2]
[1]Research Scholar Dept.of.CSE, Chaitanya Deemed to be University, Hanamkonda, T.S.
[2]Professor,Dept.of CSE, Chaitanya Deemed to be University Hanamkonda,T.S.

Abstract: The NER has been important in different applications like data Retrieval and Extraction, Text Summarization, Machine Translation, Question Answering (Q-A), etc. While several investigations have been carried out for NER in English, a high-accuracy tool still must be designed per the Literature Survey. This paper suggests an English Named Entities Recognition methodology using NLP algorithms called Bi-Directional Long short-term memory-based recurrent neural network (LSTM-RNN). Most English Language NER systems use detailed features and handcrafted algorithms with gazetteers. The proposed model is language-independent and has no domain-specific features or handcrafted algorithms. Also, it depends on semantic knowledge from word vectors realized by an unsupervised learning algorithm on an unannotated corpus. It achieved state-of-the-art performance in English without the use of any morphological research or without using gazetteers of any sort. A little database group of 200 sentences includes 3080 words. The features selection and generations are presented to catch the Name Entity. The proposed work is desired to forecast the Name Entity of the focus words in a sentence with high accuracy with the benefit of practical knowledge acquisition techniques.

Keywords: Named entity recognition, Natural language processing, Recurrent neural network, Long Short-Term Memory.

I. INTRODUCTION

Named Entity Recognition (NER) is one of the more significant responsibilities in NLP. It is the procedure in which appropriately named entities are exposed and sorted into a specific matrix of named entity matrices, for example, person name, sport, river, city, state, country, organization, etc. Unknown words can be handled in entity identification in several ways. For phrases that could be detected as unknown or for which, despite being present in the phrase check, a specific label has not been made, e.g., the Unk tag could be assigned to them to recognize them and solve the problem of unknown phrases in NER. [1]. The development of NER for Indian languages remainders essential due to the syntactic and semantic difficulties of Indian languages and the lack of equipment processing. NER is the primary package method for information retrieval, statistics extraction, query response, machine translation, and textual content summarization [2]. For example, what is important to the query processor is the realization of the point of the question (who, what, during, what, etc.). So, in many cases, the question operator corresponds to NE. The named entity machine can automatically extract predefined names (such as protein and DNA) from raw files in Statistics Biology texts. Although important paintings of the NER have been created over the years in English and various European languages, the admiration for Hindi has been very low even today. Indic languages have a rich morphology compared to other European languages. Three effective methods to implement NER are discussed. Using grammar modules in combination with machine learning or ontology or a hybrid method that contains both. Machine learning has been most prosperous in forecasting unknown entities, and rule-based structures provide the highest accuracy. Therefore, Hybrid NER structures are maximum green for Indic languages [3].

The main approaches for NER are language operations and machine learning methods. The linguistic approach uses the fully grammar-based forms that linguists write by hand. Machine learning-based technologies use many annotated learning statistics to acquire high-level language skills. Several ML strategies that can be used for the NER project are Hidden Markov Model (HMM) [7], Maximum Entropy Model (MaxEnt). Also, in Indian languages, many sentence structures can be classified as noun-entities (derivative/inflectional structures), etc., and those restrictions on those constructs differ from language to language, so detailed grammar has to be created with care for each language, and it is a task that takes Very time consuming and expensive.

In the present mechanism, these issues are solved through integration with a natively built NLP algorithm, utilizing coverage and ontological ontology represented via RDF and machine learning (ML) conditional random fields (CRF). RDF generated a list of desired named entities accumulated from different origins as a reference for the NLP and ML algorithm. Ten signs and symptoms of NER were considered, while the maximal dominant structures respected the three main symptoms: male or female, region, and company.

Entities named with phrases are also tagged to achieve higher statistics than text.

DNNs can offer more efficient confined temporal modeling by performing in a sliding window than acoustical structures of limited length. They can only edit the information within the window and get confused with the long-term dependency and audio image processing. In difference, the recurrent neural networks(RNNs) have curved that spread activations into the network from a last-time stage as input to make forecasts about the present time stage. These activations are contained within the inner conditions of the network, which, in regulation, can sustain temporary contextual activities over the long term. This mechanism allows RNNs to take full advantage of the contextual window of dynamic transformation within the input set registers rather than the fixed window as within the fixed size window used with FFN(feed-forward networks). Specifically, the LSTM structure, which overcomes several disadvantages of RNN modelling, is appealing for acoustic modelling.

The Profound Mastery Form seemed to have reached a dead end as the investigation progressed. Version performance optimization relies heavily on many drives and the best computing sources, including preschool strategies such as BERT and GPT-3. However, the data set is still a massive fee for most NLP posts, and the model's overall performance still falls short of realistic expectations from the program. At mid-length, the Insight version still has some issues with terrible playability and durability (hostile examples time).

This paper introduces NER functions and the modern NER model, which is entirely based on the RNN-LSTM model based on deep mastering.

II. RELATED WORK

NER structures can be effectively implemented in various social media and media log semantic processing pipelines. Amato et al. They exploited NER units to derive the times of applied standards from digital media files. These ideas have been used to extend the comprehension management device to facilitate seamless file management and retrieval in electronic authority software. Vanchi et al. They connected the NER unit to a full social network-based emergency controller. Social networks, which play a major role in the two-way exchange of information, can simultaneously take care of disaster control for emergencies. Entities, along with disaster locations and events, have been extracted from the textual content of social networks and posted as emergency signals to citizens in emergency areas.

Rudra Murthy et al. [4] This paper showed that the advantage discovered within deep system knowledge is a feasible modality for NER motion. We validated our deep knowledge structures in the NER CoNLL datasets in both English and Spanish. Our system can present effects similar to the major geometric structures prevailing in English. We got an impressive F-score of 89.27 for English while testing with versions that no longer use handcrafted functions or statistical sources. Evaluation of our devices trained on closer statistics indicates that the results are promising, with the effects indicated.

Wenxuan Li et al. [5] The paper summarizes named entity recognition optimization and introduces a set of named entity reputation rules based entirely on recurrent neural networks. The result indicates that improving the overall performance of the entity-specific popularity model with the help of enriching the variety and input method of datasets or presenting the model with huge computing resources for training is almost as efficient as possible with significant progression. Still would like another way to improve the NER version of Destiny Quest.

Yanyao Shen et al. [6] Deep neural networks pass out the nation of art within the popularity of the given entity. However, the advantages over classical strategies are the quality observed with less massive data units than with standard educational processes. As a result, deep learning information is more efficiently reduced when you have large units of general information or a huge price range to annotate the information manually. In these panels, we show that with the help of a combination of in-depth knowledge and active introduction, we can outperform classical techniques despite much lower learning records.

smegnew asemia et al. [7] This is seen in the mathematical improvement of the Anyuak language entity recognition of its type. NER is a fundamental sub-project in natural language processing, and the over-the-top mastery of the NER machine's money is due to the effectiveness of the final tasks. The Anyuak Entity Popularity difficulty is handled using an extended memory release to sort the tokens into predefined orders. METHODS: NER modeling long and fast memories of the Anyuak language used to trip and categorize phrases into 5 predefined phrases: personal, time, employer, region, and other (unnamed entity words). Since feature selection plays an essential role in the short-term memory framework, testing was performed on this chart to find a suitable peak amplitude for the Anyuak NER tagging project.

Gema de Vargas et al. [8] This paper described an improvement of a named entity reputation (NER) device to mechanically multiply indicators of tumor morphology into clinical files known as ICD-O, International Classification of Cancer Diseases, codes. This note was submitted as part of Cantemist's Language Technology Development Plan (TL Plan). This panel attempts to contribute winning NER

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6621
Article Received: 01 March 2023 Revised: 15 April 2023 Accepted: 01 May 2023

techniques targeting archives related to fitness in Spain. This is a critical challenge given the amount of research in the fitness field, printed in Spanish, and the uses it can bring to the scientific situation. Indeed, given that most NER techniques were implemented for English, these studies cannot be fully exploited.

Chopra, Joshi, and Mathur [9] have used Hidden Markov Model (HMM) for NER in Hindi. As they define it, HMM helps extend language-independent NER structures. These systems are also easy to measure and test. It received an F-scale of ninety-seven, 14% on training data on 2,343 symbols and testing on one hundred and five symbols. However, HMM does have a problem with naming bias. The study duration and body of evidence used in [9] can be limited.

Sharma and Goyal [10] Consider 29 functions consisting of context phrases, word prefixes, word suffixes, POS statistics, and lexicon lists: lists of individual names, company names, area names etc. finally, the accuracy, precision, and F-score are 72.78%, 65.82%, and 70.45%, respectively.

In his paper, Sinha [11] It also mentions the need to derive from an applicable set. Then, a related group was created by categorizing them into three classes, namely: NEP, NNE), and OTH. The combined CRF and rules-based output, where the output is primarily OR-ed, provides an final F-score is 71%.

N. V. Patil et al. [12] proposed an unsupervised method for extracting named entities from biomedical text. Rather than observing, your transcript fits into terminology, group stats (e.g. reverse file frequency and context vectors), and superficial syntactic experience (e.g., chopping up names). Experiments on two major biomedical data sets show his unmoderated approach's efficacy and generalizability.

III. PROPOSED METHODOLOGY

NER is a classic NLP problem identifying and classifying named entities from a text component to a predefined list of classifications.

This paper suggests leveraging the huge unlabelled data avail in this area. The Bi-Directional RNN-LSTM model is usually prepared to learn every word's recurrent and embedding layers. The embedding layer normally carries a lot of data to create useful embeddings. We develop a two-phase procedure to employ the unlabelled data.

In this proposed work, the first step includes the four stages stemming, tokenization, part of speech (POS) tagging, and chunking. The suggested method contains the procedure up to chunking so that the ambiguity of tagging stage is determined. The Named entities (NE) contain better than one token, demanding better processing after the POS tag. The NE is determined from the chunked parse tree. The identical stages are described below.

Tokenization

In this process, it is to separate words from particular texts. These texts are the primary tools for additional processing. In other texts, unique qualities may be noticed and must be removed. The result of this action is presented in the subsequent instance:

$Data: Listen\ carefully, Friends, Romans, Countrymen;$

$Yield: \{lend\}\{me\}\{your\}\{ears\}\ \{Friends\}\{Romans\}$

$\{Countrymen\}$

These tokens are usually more or less alluded to as phrases or words, but sometimes a type/token separation is necessary. A token is an example of the development of characters in a given record that can be grouped as a significant semantic unit for tracking. One type is the elegance of all the boxes that contain a compared person. The term is a type (possibly flattened) associated with an infrared device declaration reference.

Stemming and lemmatization

Stemming is a challenge that discovers the basis of each sentence, respectively. The root is legitimately applied as an operator matching a root phrase to vacate all attachments. Stemming also improves the rendering of many highlights, such as lexicon examples. The derivation is, as important as the derivation, but the crop is a lemma, not a root. Lemma is the kind of critical clause of a sentence that is regularly used as a clause of reference for a sentence. The help of Lan24 Features Local Highlights taxonomy results in the significance of derivation and lemmatization. For deeply inflectional languages such as Czech, derivation or lemmatization may be closely an absolute necessity if decreasing excessive variety in diverse word constructions is required.

Lemmatization combines different curved kinds of a word so that they can be divided as a single component. A llama is like a llama, but it carries the status of words. So, match the sentences that have equal importance to one word. Content pre-processing includes both Stemming and Stemming. Generally, people find these two words related. Some pleasure these two as identical. For all matters measured; Lemmatization is preferred over Stemming because Lemmatization does a morphologic check of words.

Part of speech and morphology

Morphological tags are a useful thing. In general, NE is often expressive of phrases and numbers. Other types, such as

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6621
***Article Received**: 01 March 2023 **Revised**: 15 April 2023 **Accepted**: 01 May 2023*

relational words, appear less frequently and, like some action words, are rarer. In arcuate dialects, morphemes also allow us to distinguish between consecutive words in similar cases, improving the identification of NE. It is a procedure of converting from sentence to systems: word abstract, group abstract (where each group has a structure (word, label)). If there should be a predicate, the tag is a syntactic procedure tag and implies whether the phrase is an aspect, signifier, transfer clause, etc.

In the historic situation of the corpus, the syntactic marking structure (POS noun, PoS tags, or POST), also known as an etymological noun or word group clarification, is the path toward word expansion in the book (corpus) in contrast to its specific bit of dissertation, subject to its description and excitatory state, i.e., the flirtation of associated and linked words in an expression, sentence, or entrance. A type of mitigation of this is often said for young children by clearly checking phrases such as subjects, words of interest, descriptors, modifiers of action words, etc.

Chunking and PoS Tagging

Segmentation is a procedure for extracting sentences from unstructured data, which resources analysing a sentence to understand the components (collections of nouns, verbs, groups of procedural sentences, etc.). However, it does not indicate its internal structure or its mission in the initial sentence. Eliminate peak point-of-sale labeling. Uses POS tags as information and provides snippets as performance. Paraphrasing can break down sentences into terms that can be more useful than single phrases, leading to great results. Hashing is great when you want to ignore content information such as versions, regions, and character names.

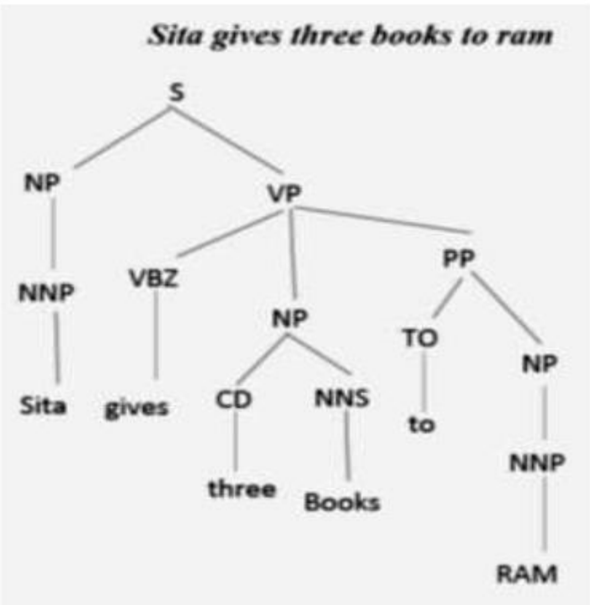

Fig.1 Tagged and Chunked Parse Tree

As shown in fig.1, “ Sita gives three books to Ram” sentence is stressed and fragmented. Note that the sentence is categorized into 2 halves. The noun word and the spoken word. Verb words and noun words can create noun phrases repeatedly. Behind naming the names, the components of the parse tree are segmented. The NEs are found in the name branches within the division model. By mapping against the current dictionary, named entities can be identified more precisely.

POS Tagging: POS tagging perfectly differentiates phrases from the right part of the grammar. The part of speech describes how the phrase is utilized in a sentence. Nouns, pronouns, adjectives, verbs, prepositions, interruptions, and conjunctions are the eight main mechanisms of speech. POS tags are a supervised learning method that uses features with prev, next word, first capital letter, and many more features. NLTK has the advantage of POS tags and works after the coding method. In this scenario, the IOB labeling method can be used. The B-chunk type is the prefix that precedes the way the tags mark the beginning of a clip. The type I-chew is the prefix indicating that it is inside a piece. O is the rating showing that the token does not belong to any part.

As shown in Figure 2, we utilized models based entirely on deep learning. We initialize the inline layers with sentence vectors for each word. Then we teach the community to quit smoking with the disaggregated data. As the various processes have shown, very good conditioning is crucial for learning the correct modifications and teaching faster (Sutskever et al., 2013). We use this method of using phrase vectors to counteract the lack of tagged records. The concept after this is that model can need vastly lower statistics for closeness and can give much more useful outcomes than when weddings are initialized randomly.

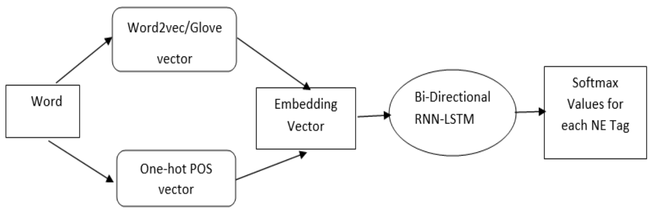

Fig.2 Embedding layers with the word vectors for every word

Bi-Directional LSTM -RNN Model:

The deep learning-based Long short-term memory (LSTM) and Conventional RNN methods have been effectively used in numerous group and group prediction tasks. In proposed language modeling, traditional RNNs have significantly reduced confusion compared to modern n-gram models, and the proposed LSTM-RNN approach has demonstrated progress over traditional LM RNNs. LSTM measures have performed better than RNNs in context-free languages.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6621
***Article Received**: 01 March 2023 **Revised**: 15 April 2023 **Accepted**: 01 May 2023*

Bidirectional LSTM (B-LSTM) networks that sequence input in both approaches have been suggested to determine the current input for phonemic tagging of sound frames in the TIMIT speech database. The Bi-LSTM networks associated with the Connectionist Temporal Classification (CTC) layer and transferred from unpartitioned succession registers have been demonstrated to exceed an instrument based on the existing HMM model for online and offline script reputation. Similar approaches with deep learning-based Bi-LSTM networks have been suggested to achieve the popularity of script-based speech. BLSTM networks have also been suggested for phonological forecast in a multi-stream framework for continuous conversational speech recognition. In terms of constructs, after realizing acoustic modelling DNN, deep RNN BLSTM combined with CTC output layer and phone sequence prediction RNN transducer has been given to achieve popular accuracy for modern phones based on TIMIT data.

PROPOSED WORK ARCHITECTURE

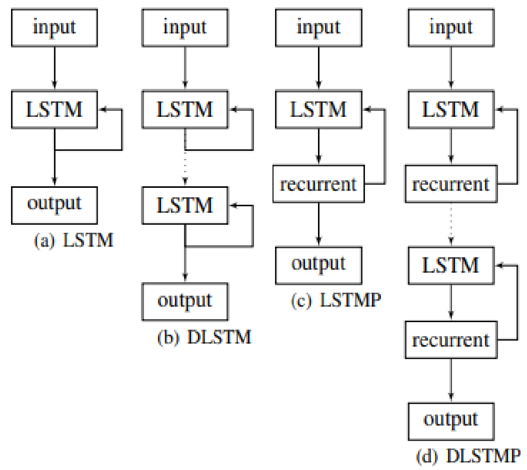

Fig.3 LSTM -RNN Architecture

Network Architecture

The proposed work of LSTM-RNN architecture is illustrated in figure 3. We prepared deep neural networks, including single or recurrent layers, because the classified data group was small. We have an embedding layer followed through one or more layers as the target within the investigations followed via the soft maximal layer. We tested three forms of iterative layers: RNN, LSTM, and finally, Bidirectional LSTM to notice which would be more appropriate for the NER scheme. The embedding layer initializes the word vector sequence and the single active vector pointing to its POS label. The POS tagging assignment is usually a helpful component for object recognition, so it has become dedicated. This speculation was validated with an advanced accuracy of 3 to 4% when containing POS tags in the embedding. This design was maximized end-to-end, utilizing Adam's optimizer (Kingma et al., 2015) and batch length 128 using a dropout layer with a dropout rate of 0.5 behind every cyclic layer. We have employed dropout (Srivastava et al., 2014) to decrease overfitting in our standards and allow the model to generalize well from the data. The main concept in dropouts is to randomly drop gadgets with their neural network associations in the context of education.

Generating Word Embeddings for English

Procedures based entirely on Word2Vec use the idea that words that occur in a similar context are comparable. Therefore, they can be grouped. There are modes added by: (Mikolov et al., 2013) CBOW and Skip-gram. The latter works better in English corporals for various tasks and is, therefore, more generalizable. Therefore, we use the aggregate approach based on the derivation scheme.

GloVe has become the latest technology for producing bulk vectors, similar to the Skip-gram-based model. Training modulations with closed window context using co-diffusion matrices. The GloVe representative is based on non-zero entrances of the joint global prevalence matrix for all words within a group. GloVe has proven to be a sufficient and widely utilized technology that can be generalized to more than one English language commitment. We use pre-studied motifs using the above methods for English because they are widely used and very effective. Vector download hyperlinks are provided. However, we teach Hindi using the above methods and make word vectors. We begin with friendly coding of words, random initialization of their word vectors, and then train them to develop word vectors finally. We utilize the LTRC IIIT Hyderabad Corpus Hindi script for training. The stats are 385MB, and the encoding used is UTF-8 layout (the unsupervised training set contains 27 million tokens and 500,000 wondrous tokens). The matched words were prepared to prepare for Hindi phrases using a context-length window of five. The expert version is then utilized to develop word embeddings in the language.

Performance of activation function

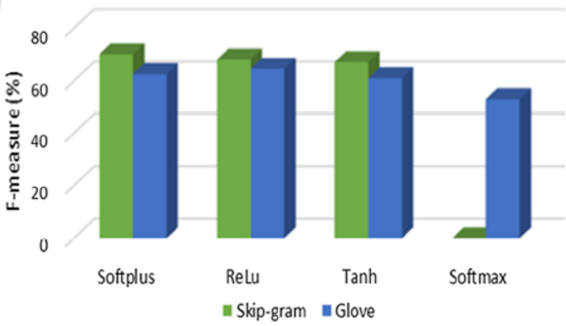

Fig.4 Performance of activation function

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6621
***Article Received**: 01 March 2023 **Revised**: 15 April 2023 **Accepted**: 01 May 2023*

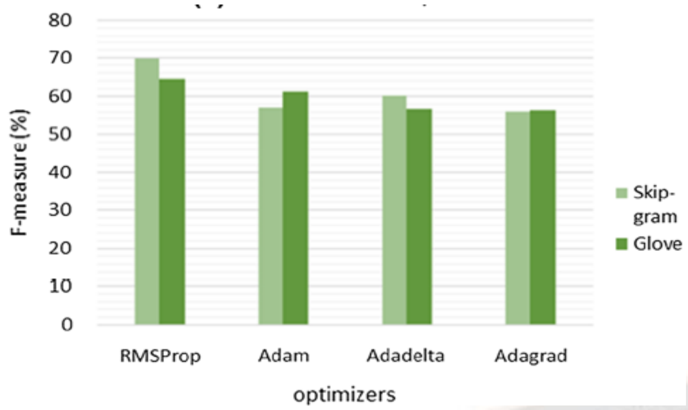

Fig.5 Performance of optimizers

In these designs, we configure various LSTM units with characteristic values of 50, 100, 150, and 200 and batch lengths, including 16, 32, 64, and 128. We refer to the performance of several hyperparameters. The system works correctly with soft plus and relu activation functions with rmsprop optimizer, as given in Figures 4 and 5.

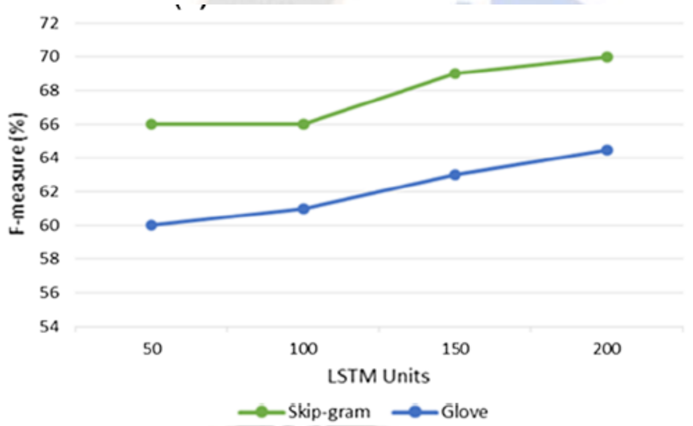

Fig.6 Performance Vs No. of LSTM units

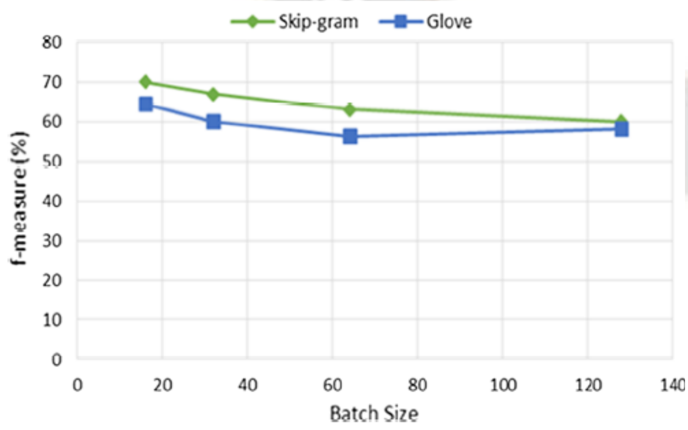

Fig.7 Performance with different hyper-parameters

The proposed approach performs for skip-gram and glove embedding better with 16 batch sizes and 200 LSTM units, as given in Fig.6 and 7.

IV. DATASET USED

A corpus has been organized for designing a baseline NER design for English data. The text data have been collected from Kaggle Dataset. This dataset includes, of course, sentences in English, but they also have related annotations for every word. The sentence in the dataset is encoded in Latin 1.

Essential info about entities:

Table.1 Named Entities with Entity tags

Named Entity	NE Tag
Geographical	geo
Organization	org
Person	per
Geopolitical	gpe
Time indicator	tim
Artifact	art
Event	eve
Natural Phenomenon	nat

$Total\ Words\ Count = 1354149$

$Target\ Data\ Column: \text{“}tag\text{”}$

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

The experiment was shown in the NLTK, a hot tool for language processing investigators. The toolkit offers a large library and instructions for conducting experiments in multiple languages, mostly English. We compare and evaluate the performance of LSTM-RNN architectures in a large word popularity project: the Google Voice Search task. We explored the empirical consequences of using intrinsic overall performance measures, namely precision (P), recall (R), and F-score (F) for other NER models that are primarily based on deep learning. Initially, we practice all modes on top of randomly generated word vectors that provide the appropriate effects. The randomly generated sentence vectors are then swapped with Facebook's pre-learned word vectors, which severely affects the accuracy of NER structures due to the unavailability of many sentences in rapid text vectors.

VI. RESULTS

The overall performance of the proposed method was estimated using various metrics. The number one metrics employed are precision, recall, and f1 score. Furthermore, it

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6621
***Article Received**: 01 March 2023 **Revised**: 15 April 2023 **Accepted**: 01 May 2023*

will use partial normal scores instead of overall average ratings because the number of tokens fitting every entity type is unbalanced. The main difference between micro-average and common macro ratings is that micro-average ratings are calculated simultaneously for all object kinds. In contrast, micro-average ratings include the computation of metrics from both sides of each entity type after combining them with the help of proposal submission. The following methods have been tried using the training dataset and quality adjustment using the disaggregated development datasets. Then, once the most useful set of parameters was achieved, the final model performance was annotated using the training and development datasets.

Table 2: Performance Metrics of NE tags

NE Tag	Precision	Recall	F1-Score
B-per	0.85	0.82	0.84
B-org	0.73	0.75	0.73
B-Geo	0.86	0.90	0.87
B-Gpe	0.98	0.94	0.96
I-per	0.86	0.88	0.87
I-org	0.73	0.82	0.78
I-Geo	0.80	0.79	0.80
I-Gpe	0.94	0.56	0.69

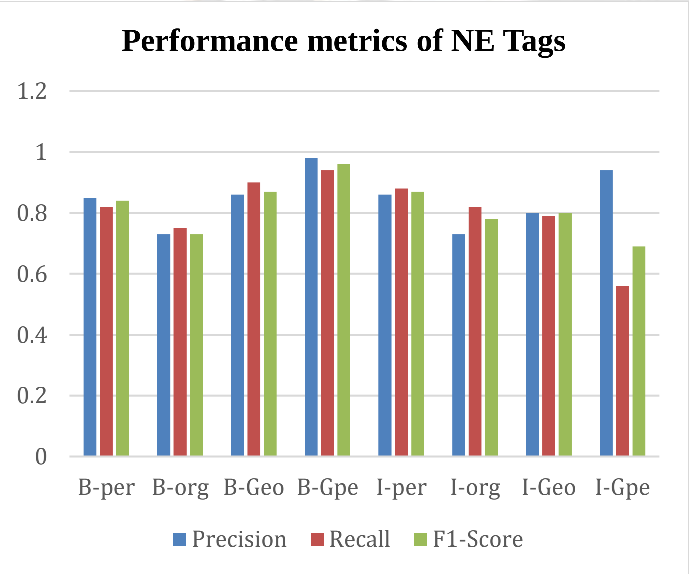

Fig.8 Performance metrics of NE Tags

Table.3 Results of several methods for NER task with skip-gram embedding

Models	Precision (%)	Recall (%)	F1-Score (%)
RNN–CNN	73	76	75
RNN–CRF	79	79	80
RNN–CNN–CRF	85	89	87
Bi-Directional LSTM-RNN	94	92	96

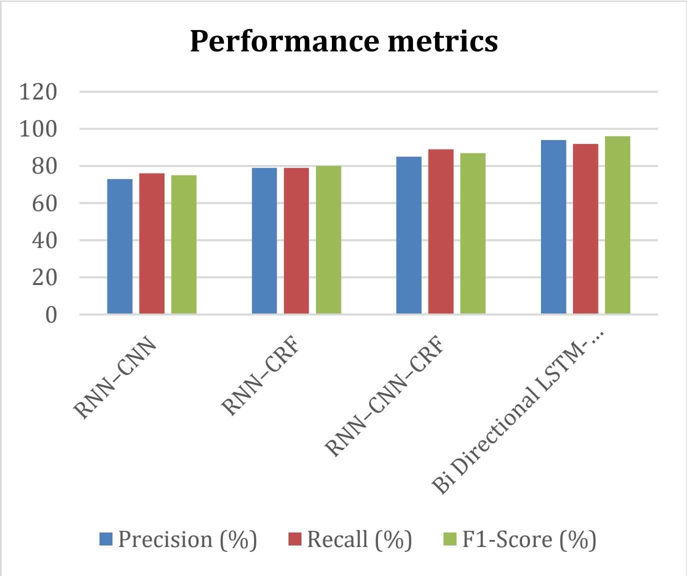

Fig.9 Various algorithms performance

As illustrated in fig.9, when the Bi-Directional LSTM model merged with the RNN, the implementation is enhanced as corresponded to other models, and it gives a Precision of 84%, Recall of 82%, and F1-score of 86%.

VII. CONCLUSION

Most current works on NER systems have been developed with a few domain names and limited script types. On the other hand, entire systems have been created for quite researched languages such as English or Hindi. Fully rule-based structures can successfully classify named entities; However, building a specific set of rules for the unrestricted domain is almost impossible. We confirm that the deep learning-based bi-directional LSTM-RNN model attains the current performance of large-scale acoustic modeling. The suggested deep learning-based Bi-directional LSTM-RNN structure is superior to standard LSTM networks and DNNs. Moreover, it uses model parameters efficiently by influencing the computational efficiency required to form a large training network. We also demonstrate for the first time that LSTM-RNN methods can be rapidly trained using the learning provided by ASGD. The release performance compared to traditional models improved performance.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6621
***Article Received**: 01 March 2023 **Revised**: 15 April 2023 **Accepted**: 01 May 2023*

REFERENCES

[1] Blythe, and A Akbik, 2018, "Contextual string embeddings for sequence labeling", pp. 1638-1649.
[2] Q Z Sheng and Zhang W E, 2020, "Adversarial attacks on deep-learning models in natural language processing: A survey[J]", pp.1-41.
[3] Q Yang & Y Zhang, 2020, 2020, "Entity enhanced BERT pre-training for Chinese NER", pp. 6384-6396.
[4] Ballesteros M, S Subramanian, 2016, "Neural Architectures for Named Entity Recognition", pp. 260-270
[5] A. Mohamed and N. Jaitly, 2013, "Hybrid speech recognition with deep bidirectional LSTM,", IEEE, 2013, pp. 273–278.
[6] Mathur I and Joshi N, 2016, "Named entity recognition in Hindi using hidden Markov model", IEEE, pp 581–586.
[7] Morwal S and D Chopra, 2016, "Hindi named entity recognition by aggregating rule-based heuristics and hidden Markov model", pp. 43–52.
[8] Coope S and Bachrach Y, 2018, "Named entity recognition with parallel recurrent neural networks" , pp 69–74.
[9] Bhattacharyya Pushpak and V Rudra Murthy, "A Deep Learning Solution to Named Entity Recognition"
[10] asemia smegnew, Gardie Birhanu, 2022, "Anyuak Language Named Entity Recognition Using Deep Learning Approach", 2998-3006.
[11] Yang J and Y Zhang, 2018, "Chinese NER using lattice LSTM". In: ACL
[12] "Workshop on NER for South and South East Asian Languages", 2020, http://ltrc.iiit.ac.in/ner-ssea-08/.