IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 5, June 2023

Advancements in Named Entity Recognition using Deep Learning Techniques: A Comprehensive Study on Emerging Trends

Shalini Sharma, Reetesh Kumar Srivastava, Dr. Piyush Pratap Singh

School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, India

Abstract*: Because of the amount of computer-processable text material and processing hardware, deep learning architecture has been the focus of knowledge projects. The task of named entity recognition in natural language processing is no exception. Because of the volume of text material that computers can process and their processing capability, knowledge projects are now focusing on deep learning architecture. Named entity recognition, is a frequently occurring task in natural language processing. Deep learning models, particularly recurrent neural net- works (RNNs) and their variations such as gated recurrent units (GRUs) and long short-term memory (LSTM) , have revolutionised natural language pro- cessing (NER) by capturing intricate contextual connections. These models can accurately forecast the borders and types of named entities because they can effectively capture the sequential and hierarchical patterns inside text. However, using deep learning in NER is a challenging task in itself. Annotated training data is critical for building reliable models, but getting labelled data for all item kinds and domains can be time-consuming. To solve this difficulty, transfer learn- ing and domain adaptation strategies have arisen, using pre-trained models and adapting them to various domains or target tasks. This paper looks at recent deep learning methods for NER and how they evolved from older linear learning approaches. It also analyses the status of activities that are either upstream or downstream of NER, such as sequence tagging and entity linking, among other things.*

Keywords: NER, Sequence tagging, LSTM, RNN.

I. INTRODUCTION

Named Entity Recognition (NER) is a NLP task that is used in identifying and classifying named entities in text . Named entities refer to specific types of words or phrases that represent entities such as persons, locations, organizations, dates, and other entities of interest. Named Entity Recognition(NER) is the process of locating a word or a phrase that references a specific entity within a text. In various NLP technologies, NER plays a major role in information retrieval and it also acts as a function for various techniques like Data Extraction, Answering Questions(Chatbot), Text Summarization, Topic modelling etc [1]. According to the study, the term "Named" refers to entities where the relationship can be represented by one or more predefined IDs (Fig 1 shows how the NER model classifies and assigns identifiers). Designators are specific names, like those of biological beings and objects, as well as natural kind terms. Given the various interpretations of NEs, the scientific community has agreed on the categories of NEs that should be recognised [2]. We usually categorise NEs into generic NEs (like people, organization or place) and domain-specific NEs [3].

Nearly four well-known methods, including rule-based methods, are utilized in named entity recognition.

- **Rule-based Approaches:** These approaches detect and classify named items by applying specified rules or patterns. Typically, these rules are created and based on language patterns, regular expressions, or gazetteers (lists of known entities). While rule-based techniques can be useful in certain areas or languages, they frequently fail to handle complex or ambiguous circumstances[4].
- **Statistical Models:** Statistical models such as Hidden Markov Models (HMMs) and Conditional Random Fields (CRFs) have been widely used in NER. Machine learning techniques are used in

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

these models to learn from labelled training data, capturing statistical patterns and connections between words and entity labels. These models frequently employ features such as part-of-speech tags, word context, and word shape [5].

- **Machine Learning Approaches:** Machine learning techniques, such as Support Vector Machines (SVMs) and decision trees, have been used in NER. These approaches entail feature engineering, in which relevant features are manually produced and fed into the learning process. Word embeddings, grammatical information, and contextual features are examples of these features. These approaches, however, are strongly reliant on the quality of feature engineering.

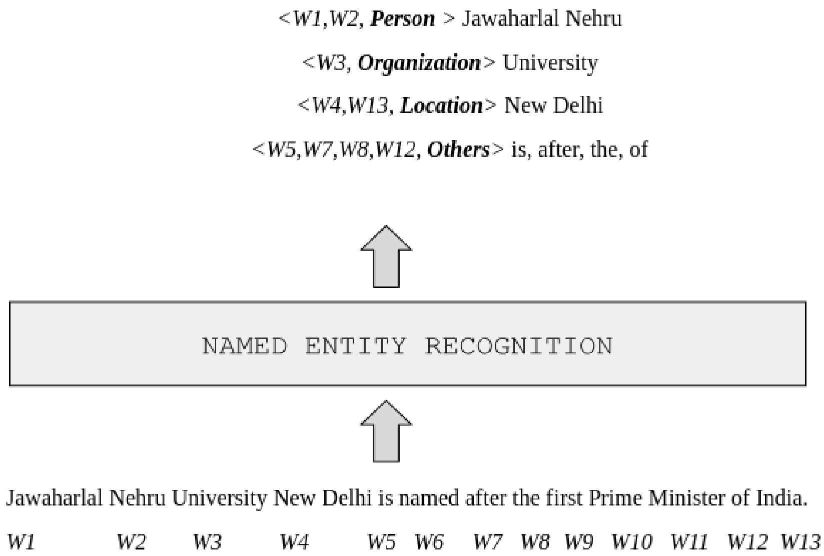

Fig. 1 Identification and Classification of Text in NER

Deep Learning Methods: Deep learning methods, like recurrent neural networks (RNNs) and its derivatives (LSTMs, GRUs), have attained cutting-edge performance in NER. These models make use of text's capacity to capture sequential and contextual information. Deep learning models for NER frequently use word embeddings to represent words, and they can be improved with approaches such as attention mechanisms and bidirectional architecture to more accurately represent dependencies and context [3].

II. LITERATURE REVIEW

Deep learning models have been showing exceptional performance in NER throughout the years, using their capacity to capture complicated contextual connections and build meaningful representations automatically.

2.1 Methodologies and Architectures

RNNs (Recurrent Neural Networks): To capture sequential information in text, early techniques used RNN-based models like Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM). These models effectively modelled word dependencies and produced competitive performance in NER tasks. In recent years, recurrent neural networks (RNNs) have gained a lot of interest. As a result of its performance in different disciplines, it has sparked tremendous attention. With minimum func- tional engineering, there is just one strong analysis of RNN-based NER frames. For the past few decades, a significant amount of research has been conducted to imple- ment RNN to NER, with the goal of continuously improving the effectiveness of the most contemporary[6].

Models Based on Transformers: With the introduction of the Transformer architec- ture, attention mechanisms have emerged as a critical component in NER models. Transformers, such as BERT (Bidirectional Encoder Representations from Trans- formers), have proven outstanding performance by taking into account word context and capturing global dependencies[7].

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

Hybrid Architectures: Several studies have been conducted to investigate hybrid architectures that incorporate the strengths of both the recurrent and transformer models. These hybrid models combine the advantages of both approaches to increase performance in NER tasks.

2.2 Datasets and Resources

- CoNLL- It is a shared task datasets, such as CoNLL-2003 and CoNLL-2002, have been frequently utilised as NER research standards. These datasets contain labelled examples of named entities from a variety of fields, including as news articles and scientific literature[8].
- OntoNotes - It is a large-scale dataset that encompasses a variety of genres, topics, and languages. It has been widely employed in the training and evaluation of NER models[9].
- Domain-Specific Datasets- To address the challenges of NER in specialised domains such as biomedical, legal, or social media texts, many researchers have concentrated on developing domain-specific datasets[10].

2.3 Performance Measures

Precision, recall, and F1-score are regularly used measures to assess the performance of NER models. Precision measures the accuracy of the projected named entity, while recall measures the coverage of the true named entity, and F1-score offers a balanced measure of both criteria.

In addition to token-level assessment, entity-level evaluation considers the entire named entity spans. This rating takes into account whether the predicted entity boundaries match the ground truth, providing an additional evaluation of NER performance[11].

2.4 Challenges

- Data Scarcity: Labelled training data for NER can be scarce, particularly in certain areas. To solve this difficulty, active research is focusing on strategies such as transfer learning, domain adaptation, and semi-supervised learning.
- Handling Ambiguity and Out-of-Vocabulary terms: Deep learning models have difficulty dealing with ambiguous entity mentions and out-of-vocabulary terms. Researchers are looking into ways to improve model resilience and generalisation.
- Multilingual NER: Due to differences in language structures and resources, expanding NER to many languages creates hurdles. Deep learning models are being used in research to establish successful solutions for multilingual NER[12].

III. DEEP LEARNING BASED NER

Deep learning is a branch of machine learning that consists of numerous processing layers that learn data representations at various degrees of abstraction. Artificial neural networks comprising forward and backward passes are common layers. The forward pass computes a weighted sum of the inputs from the preceding layer and runs the result through a non-linear function. The backward pass computes the gradient of an objective function with respect to the weights of a multi-layer stack of modules using the chain rule of derivatives. Deep learning's primary benefit is the ability of representation learning and semantic synthesis offered by both vector representation and neural processing.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

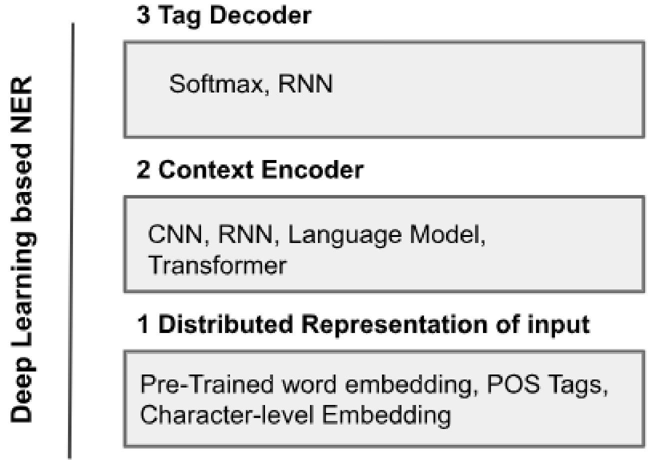

Fig. 2 Deep Learning based NER model consist of Distributed representation, Context Encoder and Tag Decoder

3.1 RNN

Recurrent neural network (RNN) can be used to successfully complete named entity recognition (NER) tasks. In NER, named entities—including people, groups, places, and other unique entities—are located and categorised inside a text. RNNs are partic- ularly well suited to NER tasks involving context analysis and relationships between words because of their capacity to grasp sequential dependencies on data. A high-level description of how an RNN may be applied to named entity recognition is given below:

- **Input Encoding:** First, the input text needs to be encoded in a language that the RNN can comprehend. Words are represented as numerical vectors in word embeddings (like Word2Vec or GloVe). As a result, the words in the input sequence are converted into the relevant word embedding vectors.
- **Sequence Modelling:** RNNs analyse encrypted input sequences one word at a time, keeping a secret internal state track of preceding words. At each time step, RNNs modify their hidden states according to the most recent embedding vector. This process is performed for each word in the input sequence

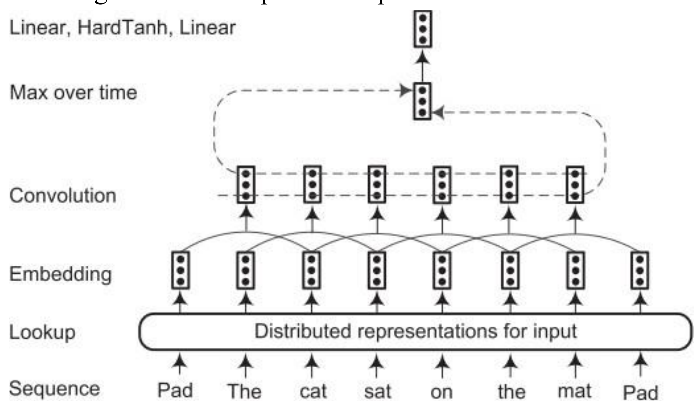

Fig. 3 RNN Based Encoder

- **Entity Classification:** The RNN's ending secret state holds assessments about the entire recommendation order after the series display process. This secret state is first provided to a classifier, to some extent a feed-forward linked system or a softmax tier, in order to predict the body labels each text. The classifier assigns a frequency distribution over the range of possible individual labels[13].
- **Training and Optimization:** Structured facts are utilised to train the model by categorising each phrase according to its appropriate item. The parameters of the RNN and the classifier are optimised

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

using techniques like backpropagation via time (BPTT) and gradient descent to reduce the appropriate loss function.

- **Prediction:** Once trained, the model may be used to draw conclusions from fresh, unexplored data. After going through the sequence modeling stage and processing the encoded input sequence, the RNN produces predictions for the entity labels of each word.

3.2 CNN

Convolutional neural networks (CNNs) can be used to successfully fulfil named entity recognition (NER) tasks. Recurrent neural networks (RNNs), which are often used for sequence modelling and are better suited for NER tasks, are able to capture local patterns and contextual information more successfully than CNNs[14].

- **Input Encoding:** Encoding the input text into numerical representations is the first stage. This often entails mapping words to word embeddings that represent the semantic links between words, such as Word2Vec or GloVe.
- **Convolution Layer:** The encoded-word embeddings are processed through one or
- more convolutional layers. Each layer contains many filters that convolve over the input. The filters extract local patterns by sliding across the input and performing element-wise multiplications followed by summing. This technique produces feature maps that highlight important local information

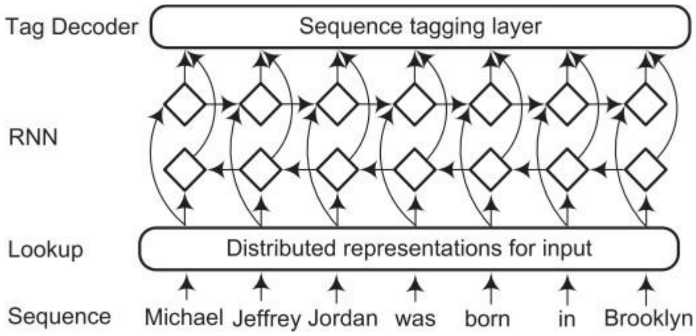

Fig. 4 CNN based encoder

- **Fully Connected Layers:** Following flattening, one or more completely linked layers are routed through the pooled feature maps. These layers take the retrieved features and analyze them further, allowing the model to learn more complex representations of the input.
- **Output Layer** The final fully linked layer is connected to the output layer, which creates predictions for each word in the input text. When utilising NER, the output layer typically uses a SoftMax activation function to provide probabilities for each named entity class (such as a person, company, location, etc.) for each word.

3.3 LSTM

The ability of LSTM (Long Short-Term Memory) networks to recognize long-distance relationships in sequential data have led to their widespread usage in named entity recognition (NER) tasks and good performance.

- **Input Encoding:** Encoding the input text into numerical representations is the first stage. Similar to other neural network techniques, word embeddings are frequently used to map words to semantic relationships between words.
- **LSTM Layer:** The encoded-word embeddings are fed into an LSTM layer. Word-
- by-word sequential input is processed by the LSTM layer while a state is kept in internal memory. The most recent word embedding is given to the LSTM cell as input, and at each time step, it updates its memory state and output based on the most recent input and the previous memory state.
- **Bi-directional LSTM:** A bi-directional LSTM is a popular method for capturing data from both past and future words. To develop this model two LSTMs are used. one of which processes the input sequence in the forward direction and the other of which processes it in the reverse manner[15].

- **Output:** The final fully linked layer is connected to the output layer, which creates redictions for each word in the input text. When utilising NER, the output layer typically uses a SoftMax activation function to provide probabilities for each named entity class (such as a person, business, place, etc.) for each word

IV. CONCLUSION

In this paper we have seen various Deep learning models and their parameters like Input Encoding, Sequence Modelling , Classification, Training and Prediction methods of three main models RNN, CNN and LSTM.

- RNNs are a deep learning model that can analyse sequential data by processing inputs one at a time while remaining hidden. RNNs may capture dependencies between words in a phrase and keep contextual information in named entity representation. Traditional RNNs, on the other hand, suffer from the vanishing gradient problem, which limits their ability to capture long-term dependencies.
- CNNs typically serve as tools for image processing, but they may also be used to process text data by considering text as a one-dimensional signal. CNNs may extract local features from fixed-size windows of words in a phrase in named entity representation. They are good at catching local patterns and can learn meaningful representations, but they may struggle with long-range relationships.
- LSTM networks are a form of RNN that was created to solve the vanishing gradient problem. By keeping a memory cell and several gates active, they may capture long- term dependencies in sequences. Because they can efficiently model context-specific data, manage sequences of variable lengths, and maintain information over long distances, LSTMs are often utilised in named entity representation tasks.

In conclusion, named entity representation has been implemented using RNNs, CNNs, and LSTMs. LSTMs and other RNNs are excellent at capturing sequential dependen- cies and contextual data. CNNs can extract regional traits and patterns from text, despite being predominantly utilised for image processing. Being a subset of RNNs, LSTMs are frequently used because they can capture both short-term and long-term dependencies. The method of processing the data is determined by the particular needs of the named entity representation task as well as the characteristics of the data itself.

REFERENCES

[1]. Mansouri, A., Affendey, L. S. & Mamat, A. Named entity recognition approaches. International Journal of Computer Science and Network Security 8, 339–344 (2008).
[2]. Nadeau, D. & Sekine, S. A survey of named entity recognition and classification. Lingvisticae Investigationes 30, 3–26 (2007).
[3]. Li, J., Sun, A., Han, J. & Li, C. A survey on deep learning for named entity recognition. IEEE Transactions on Knowledge and Data Engineering 34, 50–70 (2020).
[4]. Riaz, K. Rule-based named entity recognition in urdu, 126–135 (2010).
[5]. Ferreira, E., Balsa, J. & Branco, A. Combining rule-based and statistical methods for named entity recognition in portuguese (2007).
[6]. Lyu, C., Chen, B., Ren, Y. & Ji, D. Long short-term memory rnn for biomedical named entity recognition. BMC bioinformatics 18, 1–11 (2017).
[7]. Souza, F., Nogueira, R. & Lotufo, R. Portuguese named entity recognition using bert-crf. arXiv preprint arXiv:1909.10649 (2019).
[8]. Sang, E. F. & De Meulder, F. Introduction to the conll-2003 shared task: Language-independent named entity recognition. arXiv preprint cs/0306050 (2003).
[9]. Pradhan, S. S. et al. Ontonotes: A unified relational semantic representation, 517–526 (IEEE, 2007).
[10]. Zhang, X. et al. Domain-specific ner via retrieving correlated samples. arXiv preprint arXiv:2208.12995 (2022).
[11]. Yadav, V. & Bethard, S. A survey on recent advances in named entity recognition from deep learning models. arXiv preprint arXiv:1910.11470 (2019).

[12]. Rahimi, A., Li, Y. & Cohn, T. Massively multilingual transfer for ner. arXiv preprint arXiv:1902.00193 (2019).

[13]. Li, J. et al. Wcp-rnn: a novel rnn-based approach for bio-ner in chinese emrs: Paper id: Fc 17 25. The journal of supercomputing 76, 1450–1467 (2020).

[14]. Wang, J., Xu, W., Fu, X., Xu, G. & Wu, Y. Astral: adversarial trained lstm-cnn for named entity recognition. Knowledge-Based Systems 197, 105842 (2020).

[15]. Dong, X. et al. Transfer bi-directional lstm rnn for named entity recognition in chinese electronic medical records, 1–4 (IEEE, 2017).