ISSN 2278-3091
Volume 8, No.6, November – December 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse150862019.pdf
https://doi.org/10.30534/ijatcse/2019/150862019

Sequence Labeling Using Deep Neural Nets

N V Nikhila[1], P Himabindu[2], R Subbarao[3], S Sagar imambi[4]
[1]K L Deemed To Be University, India, nadellanikhila22@gmail.com
[2]K L Deemed To Be University, India, himabindupuritipati1999@gmail.com
[3]K L Deemed To Be University, India, subbarao869@gmail.com
[4]K L Deemed To Be University, India, simambi@kluniversity.in

ABSTRACT

Now a day's sequence labeling has become most interesting topic in the current technical era. Sequence labeling is a type of pattern recognition task that involves the algorithmic assignment of a categorical label to each member of sequence of observed values, and also, it is treated as an independent task. By using traditional methods such as HMM and CRF we can implement sequence labeling. Both the methods take the sequence of input and learn to predict an optimal sequence of labels. These are very powerful methods, but they have not experienced the great success due to some drawbacks like lack of semantic awareness and can't handle longer sequential dependencies. So by using deep learning techniques such as recurrent neural networks, they can capture the local dependencies and find longer patterns. Real world applications where the sequence labeling can be applied are Google search Engine. Where in the search box if we type some words automatically Google will suggest some sentences or words which makes our work easier.

Key words: Bi-Lstm, CRF, HMM, RNN

1. INTRODUCTION

Sequence labeling is one of the main tasks that are focusing on Natural Language Processing over the past decade. The main aim of NLP is to convert a human language into a formal representation so that it is easy for the computers to manipulate. Some of the applications are search, information extraction, machine translation [1]. POS, Word Sense Disambiguation, NER, Word Segmentation are some of the subtasks of sequence labeling. Among these NER is the important task of NLP. The most traditional sequence labeling models which have shown high performance are linear statistical models like HMM and CRF [2] but these models highly rely on specific task resources and on hand crafted features. By using high performance approaches we can get the better results when compared to the linear statistical Models. Accuracy what we get through these linear statistical models is not accurate when compared to the deep learning. techniques. So to overcome the drawbacks in the existing techniques and to get the good results we are training and testing the datasets by using Deep learning techniques. So that we can improve the accuracy [3-5]. Deep Learning is the key Technology that is used in many exciting novel applications like Google translators like Siri and Alexa [6]. The difference between the traditional methods and the deep learning methods is both the methods take the sequence of input and learn the optimal sequence and predict the labels for the sequence, But the traditional methods like HMM simply works on the words (type of tokens), and the CRF works on the set of some features like input token or phrases. These are the very powerful methods but they have not been experienced great success due to some of the drawbacks. By using the deep neural nets, we can overcome some of the drawbacks and they have shown great power to learn latent features.

2. LITERATURE SURVEY

2.1 Sequence Labeling

Sequence labeling is a type of machine learning technique which is used for pattern recognition and allows categorizing labels for the observed values. Some of the techniques of Sequence Labeling are Parts Of Speech Tagging(POS), Named Entity Recognition(NER), Word Segmentation[13].
Initially linear statistical models used for sequence labeling,The models are HMM CRF and SVM.

2.2 HMM

HMM is a generative model which assigns the joint probability for the observations and the label sequence. Then the parameters are trained so that to maximize the joint likelihood of training sets [7]. HMM is called hidden because only the symbols emitted by the system we can see, but not the process which is undergoing between the layers or the states [10]. In HMM there will be Transition Probabilities and Emission Probabilities.

*Transition Probability***:** It is the probability of the state '*s*' appearing after observing sequences *u* and *v* in the sequence of observations.

$$Q(s|u, v)$$

*Emission Probability***:** It is the Probability of an observation *x* given that the state was *s*

$$E(x|s)$$

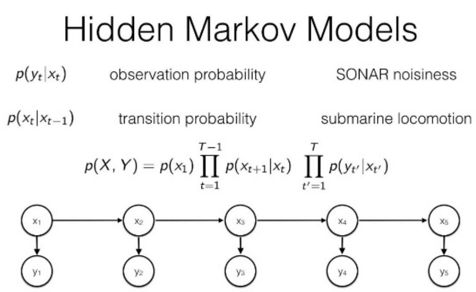

Figure 1: HMM

There are some drawbacks like here there are many unstructured parameters and they cannot express the dependencies among the hidden states and are unable to capture the higher order information. To overcome these drawbacks CRF came into existence.

2.3 CRF

This is a statistical model. This is widely used for pattern recognition[16]. This is one of the category of sequence modelling. This is undirected probabilistic graphical model. Especially this method is used in pos tagging and in the NER and also in the parsing and shallow parsing. This consists of two layers input layer, output layer. This method is used to encode the relationships between each of the observations[16].

2.4 SVM

This method is one of the traditional based approaches. This is used to separate the data by using the hyper plane. Based on the data the hyper planes may vary [11]. This can convert the low dimensional data into high dimensional data by using the kernel tricks.

3. PROPOSED METHOD

Bi-LSTM-CNN:

It is the combination of both the neural network models can use for getting better accuracy in NER. It is the hybrid model because it is the combination of both bi-directional LSTMs and CNNs[22].

Bi-LSTM: This method is a combination of 2 techniques. They are LSTM and Bi-RNN(Bi-directional Recurrent neural networks). Bi-LSTM is an advancement or special development of the Artificial Neural networks(ANN). This method came into existence because in the previous methods for larger sequence of the data the traditional methods are not suitable to solve the problems so to overcome this disadvantage this method is used[12]. And also RNN is not supportable for this Longer sequence of data. Bi-LSTM is supportable for the longer sequence of data and it is an advancement to the RNN. However, these Bi-LSTM consists of three layers Input layer , hidden layer, Output layer.

Bi-LSTM can do both forward and backward operations. So this part becomes the main advantage in this model. Long short term model is capable of storing the information for longer period of time. And also it stores the previous state information and sends to the next layer in a sequence format. If the current layer wants the previous information this BI-LSTM is used. In this BI-LSTM there will be forward hidden layer and also the backward hidden layer. Which helps us in storing the past information of the input layer. Nodes will be present in each layer of the Bi-LSTM. The nodes in the hidden layer are connected this is how the information will be stored in the layers.

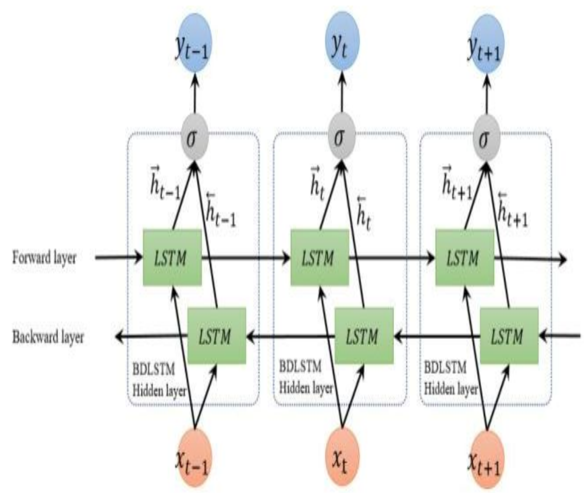

Figure 2: Bi-Lstm

CNN: This one type of artificial neural network(ANN). This method is used in pattern recognition and also used in the sequence labeling part. As this method consists of 3 layers they are Convolution layer and the fully connected layer and also the output layer [14]. This method is mainly used in the name entity recognition(NER). The main task is that it detects the words or the characters present in the text. So by using this BI-LSTM method we can correctly characterize the words and also the characters into the levels format[18].

In this CNN back propagation technique is also available. This becomes the main advantage to the CNN. After getting the result to the output layer[19]. We will check the result and compare with the actual result if the result does not match with the actual result then by using back propagation method we will update the values and we will try to achieve the better and the accurate results [15].

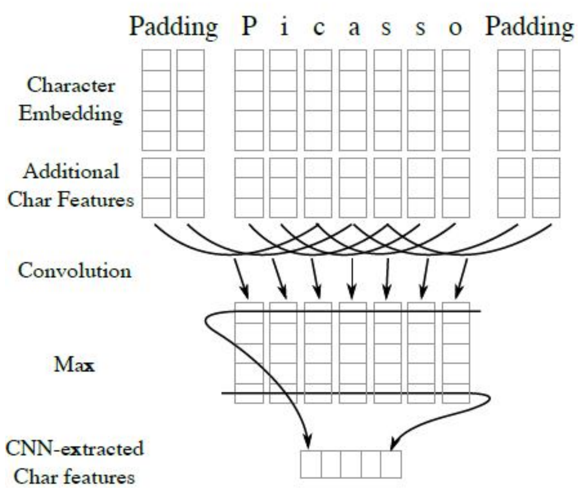

Figure 3: CNN

Fig 3 explains that the character features are extracted from each word by using CNN and the character type feature vector and character embedding are computed using look up tables after that they are concatenated and passed into the model CNN.

These models learns about both word level and character level features. This model is inspired by the work of Colobert et al.(2011)

Here the lookup tables will transform the discrete features like characters and words into contionous vector representations and then these are concatenated[23] and fed into a BiLSTM network. After that for inducing character level features use a CNN[20] and then

At the output layers it decodes the output for each category into a score.

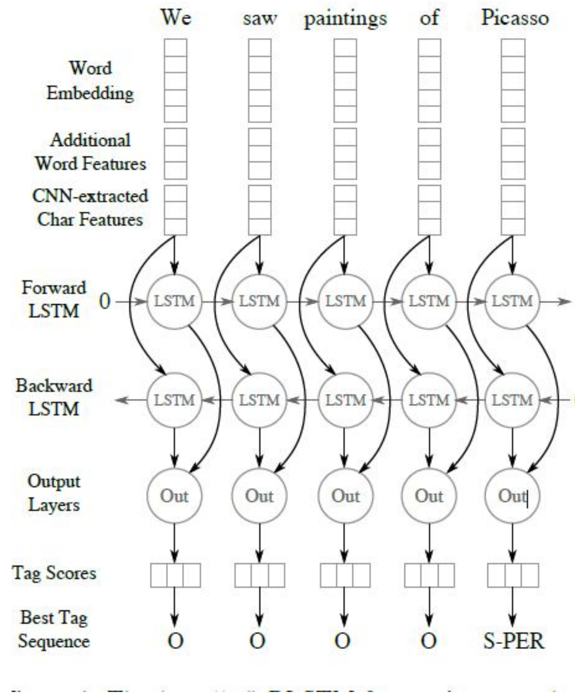

Figure 4: BI-LSTM-CNN

The above figure explains how the Bi_LSTM used for named entity recognition[21].

4. ALGORITHM

1. Reading the dataset
2. Pre-processing techniques
   - If there are Null Values, fill those values By preceding values
   - Segmentation
   - Tokenization
3. Splitting the dataset into train and test sets
4. Applying the algorithms to the trained and testing modules
5. Apply the CRF, BI-LSTM CNN on the dataset
6. Train a CRF and also the BI-LSTM CNN model using Sklearn-Crfusuite on dataset
7. Feature extraction most of the features are like word parts, simplified POS tags, lower, title, upper flags, features of nearby words
8. Evaluated the model by observing the scores of Precession, Recall, F1-Score.

The above algorithm explains about the steps involved in implementing the Bi-LSTM-CNN and CRF

5. RESULTS

The dataset used in this paper is NER dataset. It consists of 4 columns

1. Sentence#
2. Word
3. POS
4. Tag.

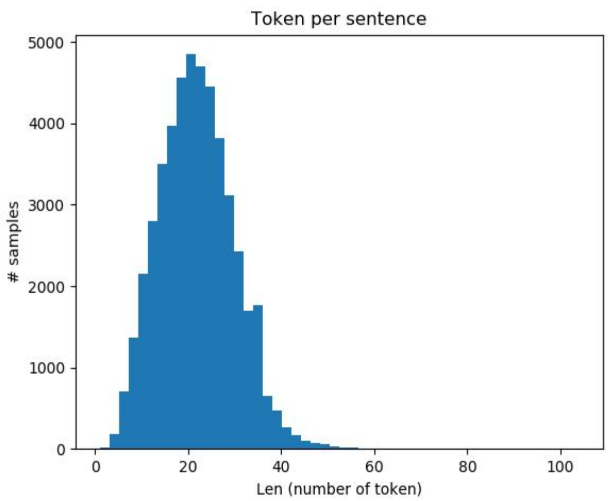

Figure 5: representation of dataset

This graph states that the samples what we have taken and also about the number of tokens which are assigned to a dataset.

After that Named Entity Recognition is performed by using Bi-LSTM-CNN which is the task of sequence labeling. Initially preprocessing of data takes place, if there are any null values it will do preprocessing and the next step is splitting of dataset into train and test sets so that by applying the CNN model on training and testing sets CNN character features are extracted and by using BiLSTM the output sequence and the performance scores are obtained.

5.1 SAMPLE INPUT

"*Stephen went to America*"

It will predict the named entities for each word by using the BiLstm Model and then gives the evaluation scores F1 Score, Recall, Precision and accuracy.

5.2 SAMPLE OUTPUT

Stephen – 'B-PER'

Went – 'O'

To –'O'

America – 'B-LOC'

It will check whether the named entities are correct or not and gives the evaluation scores.

5.3 PERFORMANCE MEASURES

True Positives(TP): These values are the correctly predicted (+Ve) values which means the actual class is yeas and the predicted class is also yes.

True Negatives(TN): These values are the correctly predicted (-Ve) classes which means the actual class is no and the predicted class also no.

False Positives(FP): These values are wrongly predicted where the actual class is No and the predicted class is yes.

False Negatives(FN): These values are wrongly predicted where the actual class is yes and the predicted value is No.

Accuracy(Ac): It is the performance measure and it is the ratio between the correctly predicted and total number of observations.

$$Accuracy = TP + TN / TP + FP + FN + TN$$

Precision: It is the Ratio between the Correctly predicted positive observations and the total predicted positive observations.

$$Precision = TP / TP + FP$$

Recall: It is the ratio between the correctly predicted positive observations and all the observations in actual class.

$$Recall = TP / TP + FN$$

F1 Score: It is the weighted average of Precision and Recall.

$$F1\ Score = 2 * (Recall * Precision) / (Recall + Precision)$$

5.4 Comparison Table

Table 1: Comparison of results

	F1	Recall	Precision	Accuracy
SVM	0.57	0.54	0.54	0.56
CRF	0.80	0.78	0.78	78
Bi-LSTM-CNN	0.97	0.97	0.97	97

6. CONCLUSON

In this paper we have compared the two models. This comparison is done to find out the accuracy among the methods. So in this paper the techniques what we compared are Bi-LSTM CNN and CRF models. BI-LSTM CNN got more accuracy when compared to CRF. After getting the results this paper finally concludes that deep learning methods are more suitable for sequence labelling because the accuracy is more to the deep learning methods when compared to the traditional methods.

REFERENCES

[1]. R. COLLOBERT AND J. WESTON, "**A UNIFIED ARCHITECTURE FOR NATURAL LANGUAGE PROCESSING: DEEP NEURAL NETWORKS WITH MULTITASK LEARNING**," IN PROCEEDINGS OF THE 25TH INTERNATIONAL CONFERENCE ON MACHINE LEARNING. ACM, 2008

[2]. Domic Seyler, Tatiana Dembelova, Luciano Del Corro, Johannes Hoffart and Gerhard Weikum, "**A study of the Importance of External Knowledge in the Named Entity Ecognition Task**." In proceedings of 56th Annual Meeting of the Association for Computational Linguistics, 2018 https://doi.org/10.18653/v1/P18-2039

[3]. Jason P.c. Chiu, Eric Nichols, "**Named Entity Recognition with Bidirectional LSTM-CNN-CRFs**," in Transactions of the Association For Computational Linguistics, vol. 4, 2016 https://doi.org/10.1162/tacl_a_00104

[4]. Xuezhe Ma, Eduard Hovy, "**End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF**" in proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016

[5]. Shotaro Misawa, Motoki Taniguchi, Yasuhide Miura and Tomoko Ohkuma, "**Character based Bidirectional LSTM_CRF with words and characters for Japanese Named Entity Recognition**," in proceedings of the first Workshop on subword and Character Level Models in NLP, 2017

[6]. Chao Li, "**Sequence Labeling Via Deep Learning – The magic behind Extract**!" 2019

[7]. Sudha Morwal, Nursat Jahan, Deepti Chopra, "**Named Entity Recognition using Hidden Markov Model (HMM)**",International Journal on Natural Language Computing(IJNLC) Vol.1 , No. 4, December 2012 https://doi.org/10.5121/ijnlc.2012.1402

[8]. Zachary C. Lipton Charles Elkan David C. Kale Randall Wetzel "**Learning to Diagnose with LSTM Recurrent Neural Networks**" publishes as conference paper at ICLR 2016.

[9]. Yingwei Xin, Ethan Hart, Vibhuti Mahajan, Jean-David Ruvini, " **Learning Better Internal Structure of words for Sequence Labeling**", proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018.

[10]. Yasemin Altum, Ionnis Tsochantaridis, Thomas Hofmann, "**Hidden Markov Support Vector Machines**", proceedings of the twentieth International conference on machine Learning (ICML-2003), 2003

[11]. S. Sathiya Keerthi, S. Sundararajan, "**CRF versus SVM-Struct for Sequence Labeling**"

[12]. Sak, Hasim/senior, Andrew / Beaufays, Francoise (2014): "**Long short-Term memory recurrent neural network architectures for large scale acoustic modeling**"

[13]. Muhammad Rifayat Samee, Mercer, Robert e: "**Improving Neural Sequence labeling using additional linguistic information**"

[14]. Hu xu, Bing Liu, Lei Shu, Philip S. Yu "**Double Embeddings and CNN based Sequence labeling for aspect Extraction**" proceedings of 56th annual meeting of the association for computational linguistics, 2018

[15]. Nam Nguyen, Yunsong Guo, "**Comparisons of Sequence labeling Algorithms and Extensions**" in proceedings of the 24th International Conference on Machine Learning, 2007

[16]. Shuai zheng, sadeep jayasumana, Bernardino Romera-Paredes, Vibhav Vineet, Zhizhong Su Dalong Du, ChangHuang Phillip H.S. Torr, Torr-Vision Group, University of Oxford Standard University Baidu Research.

[17]. Navya Krishna , Rajarajeswari P et al " **Recognition of Fake Currency Note using Convolutional Neural Networks** " International Journal of Innovative Technology and Exploring Engineering (IJITEE) , 2019 , Volume-8 Issue-5, pp: 2278-3075

[18]. Pellakuri Vidyullatha, Rajeswara Rao D., "**Modeling of supervised ADALINE neural network learning technique**", Proceedings of the 2016 2nd International Conference on Contemporary Computing and Informatics, 2016, , pp.17-22 https://doi.org/10.1109/IC3I.2016.7917928

[19]. S. Rizwana, S.Sagar Imambi "**Enhanced biomedical data modeling using unsupervised probabilistic machine learning technique**" International Journal of Recent Technology and Engineering,,2019, vol7.No 6. Pp 579-582

[20]._Banerjee D, S.Sagar Imambi. " **Opinion mining for drug reviews** " International Journal of Innovative Technology and Exploring Engineering. Vol 8 No.7, pp. 1314- pp.13-18

[21]. Lakshmi Prasanna, P., Rajeswara Rao, D., "**Probabilistic Recurrent Neural network for topic modeling** ",International Journal of Innovative Technology and Exploring Engineering , 2019 ,8(4), pp. 165-168

[22]. Greeshma, L., and G. Pradeepini. "**Mining Maximal Efficient Closed Itemsets without Any Redundancy**." Information Systems Design and Intelligent Applications. Springer, New Delhi, 2016. 339-347.

[23]. R. Bhimanpallewar, M. R. Narasinagrao, "**A Machine Learning Approach to Assess Crop Specific Suitability for Small /Marginal Scale Croplands**" International Journal of Applied Engineering Research 2017 Vol.12, NO.23, ppno.13966-13973.

[24]. Maganti syamala, N. J. Nalini, "**A Deep Analysis on Aspect based Sentiment Text Classification Approaches.**" International Journal Of Advanced Trends In Computer Science And Engineering(IJATCSE) Volume 8, No.5, September - October 2019.
https://doi.org/10.30534/ijatcse/2019/01852019

[25]. Sheetal S. Pandya, Nilesh B. Kalani, "**Review on Text Sequence Processing with use of different Deep Neural Network Model"** International Journal of Advanced Trends in Computer Science and Engineering(IJATCSE) Volume 8, No.5, September - October 2019
https://doi.org/10.30534/ijatcse/2019/56852019