CSCE 5290 - Natural Language Processing
Project Proposal

Title: Named Entity Recognition using BiLSTM

Team members:

Venkata Saran Praneeth Koduru (11440082)

Md Farhad Mokter (11336535)

Goals and Objectives:

Named Entity Recognition (NER) in textual documents is an essential application of Natural Language processing. The application of NER require a set of informative features in terms of linguistic patterns upon well engineering based on domain engineering. We intend to apply NER model using BI-LSTM in this project.

Significance:

By using this model, named entities can be identified and can be tagged with the appropriate label. The project can be used for different domain i.e. clinical entity recognition for clinical data analysis and data mining.

Objective:

The objective of this model is to develop and implement a NER model to identify the named entities automatically. IT can also be seen as a sequence tagging task where we intend to assign a unique label to each tokens. We will do all necessary pre-processing to extract the tokens to feed the model.

Dataset and features:

We would use one publicly available datasets : CONLL 2003 (English) for our project. We will try make some modification to the dataset for our task by doing some pre-processing. We intend to tag a given

sequence by using our model and assign labels like Person (P), Location(L), Origination (O). We would also try to assign undecided words into a Miscellaneous (M) label and create a “Not sure” label for such complex words. We aim to implement a pre-trained model to assign proper or accurate label to these “Not-sure” words for sake of improving our proposed project accuracy. We would also employ new ideas and algorithms upon recommendations and results.

References:

1. Mayhew, Stephen, Gupta Nitish, and Dan Roth. "Robust named entity recognition with truecasing pretraining." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. No. 05. 2020.
2. Thanh Hai Dang, Hoang-Quynh Le, Trang M Nguyen, Sinh T Vu, D3NER: biomedical named entity recognition using CRF-biLSTM improved with fine-tuned embeddings of various linguistic information, *Bioinformatics*, Volume 34, Issue 20, 15 October 2018, Pages 3539– 3546, https://doi.org/10.1093/bioinformatics/bty356
3. G. Yang and H. Xu, "Named Entity Recognition via Interlayer Attention Residual LSTM," 2022 International Joint Conference on Neural Networks (IJCNN), 2022, pp. 01-08, doi: 10.1109/IJCNN55064.2022.9892493.

GitHub :- https://github.com/praneethk6795/NER-Using-BiLSTM