Sequence Labeling: Named Entity Recognition

Supervisor: Oier Lopez de Lacalle & Eneko Agirre

1 Description

In this project you will build a deep learning model that labels words in sequences. In particular, we will focus on Named Entity Recognition, in which the task is to find the entity type (e.g. PSERSON, LOCATION, ORGANIZATION, etc.) of each word in the sequence. The sencte below shows an output of the system where multiple entities are recognized and classified:

*[*PER *Wolff], currently a journalist in [*LOC *Argentina], played with [*PER *Del Bosque] in the final years of the seventies in [*ORG *Real Madrid].*

The project can be approached that would require the use of deep learning architectures such as CNNs and LSTMs. Implemented model will be evaluated according to the standard metrics used in the field.

2 Objectives

The objectives are flexible depending on the desired level of difficulty:

- D1: Implement a LSTM based sequence labeler. Optionally, basic model can be extended by addng a CRF layer [1].
- D2: Implement a sequence labeler that combines CNN and LSTM in the same model [2].

Objectives will be adjusted with the supervisor, and the final mark will depend on what is agreed.

3 Materials

We will provide some pointers to be able to code the parts the sequence labeler. We will provide training and testing datasets. Please contact Oier to check the details of the objectives, helper codes, as well as the datasets (oier.lopezdelacalle at ehu.eus).

References

[1] Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, Chris Dyer. "Neural Architectures for Named Entity Recognition". Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 260–270. 2016

[2] Chiu, Jason P.C. and Nichols, Eric. "Named Entity Recognition with Bidirectional LSTM-CNNs". In Transactions of the Association for Computational Linguistics, pp. 357–370. 2016.