COMPARING CONDITIONAL RANDOM FIELDS AND LSTM NETWORKS FOR NAMED ENTITY RECOGNITION

Josef Gugglberger

supervised by Dipl.-Ing.Clemens Sauerwein, PhD

Masterseminar

leopold-franzens-universität innsbruck

December 2019

Abstract

Contents

Chapter 1

Introduction

With the rise of deep learning in the last few years (from 2015 onwards), the methods to perform named entity recognition (NER) have changed in research. The traditional machine learning approaches like Hidden Markov Models and Conditional Random Fields (CRF) were replaced by a special type of Recurrent Neural Networks, so-called Long-Short-Term-Memory networks (LSTM). The first approach to apply LSTM networks for named entity recognition was done already in the year 2003 in the paper [3]. However, due to the lack of computational power back then, the performance was far behind state-of-the-art methods. In 2015, Huang et al. [5] showed various LSTM network that reached state-of-the-art performance. Also, the currently best performing NER system [2] on the CoNLL 2003 dataset makes use of some kind of LSTM network.

Aim of this work is to perform a practical study, comparing the implementation and the performance of a machine learning approach, namely CRFs, and a deep learning approach implementing a LSTM network.

The remainder of this thesis is structured as follows: Chapter 2 is about related work and some background information on named entity recognition, CRFs and LSTM networks. Next, Chapter 3 describes the details of the two implementations of the NER system. Chapter 4 compares the results generated by the previous chapter.

Finally, Chapter 5 concludes and give an outlook what can be done in future works.

Chapter 2

Background & Related Work

This chapter presents basic background information of the topics NER (Section 2.1), CRF (Section 2.2) and LSTM networks (Section 2.3). The last section, Section 2.4, summarizes the recent work that was done in the area of NER.

2.1 Named Entity Recognition

Named Entity Recognition is the task of locating and classifying named entities in unstructured text. A named entity is classified into a predefined set of categories, which can be names of people, locations, companies, etc. A named entity describes something physical (person, location, etc.), and a *mention* is a phrase in the text that refers to that entity with a name [8]. For example, the terms *U.S.A* and *the United States of America* are two different phrases that refer to the same named entity.

The categories in which the named entities are classified depends strongly on the domain where the NER system will be deployed, but the categories *person*, *location*, *organization* and *miscellaneous* are often used in research, for example in [9].

According to AclWeb [1], the current state of the art approach on the CoNLL-2003 [9] dataset reaches a F1 score of 93.09 %.

2.2 Conditional Random Fields

A *Conditional Random Field*, initially proposed in [6], is a discriminative probabilistic classifier. In contrast to a discrete classifier (like the naive Bayes classifier), it makes its prediction not just based on the input sample, but also based on the context of the input sample. This is an important feature in the domain of natural language processing, especially in part-of-speech tagging and named entity recognition.

To model the context of a sample we have to define feature functions, which have to be defined in a way that expresses some characteristic that is present in the training data and we want our model to hold. A feature function has to form:

$$f(y_t, y_{t-1}, x_t) \tag{2.1}$$

where y_t is the current label, y_{t-1} is the previous label and x_t are the samples from the input sequence that the feature function needs to calculate the feature. The Conditional Random Field, as defined in [8], is then:

$$p(y|x) = \frac{1}{Z(x)} \prod_{t=1}^{T} exp(\sum_{k=1}^{K} \theta_k f_k(y_t, y_{t-1}, x_t)) \tag{2.2}$$

where $Z(x)$ is an normalization function:

$$Z(x) = \sum_{y} \prod_{t=1}^{T} exp(\sum_{k=1}^{K} \theta_k f_k(y_t, y_{t-1}, x_t)) \tag{2.3}$$

where X and Y are random vectors and θ is a parameter vector. The values of the parameter vector θ have to be learned from the training data, usually via Maximum Likelihood Estimation.

2.3 LSTM Networks

A LSTM network is a special type of Recurrent Neural Network (RNN). In contrast to a feed-forward neural network, like for example Convolutional Neural Network, where the date flows only in one direction (from input to output), the data can flow in cycles

in a RNN. This property makes a RNN able to memorize inputs, which makes them suitable for sequence labeling tasks. RNNs have been considered difficult to train a long time, especially when they are trained to recognise long-term dependencies. The problems that arise in such cases is called the *vanishing* and *exploding* gradient problem [7].

LSMT networks, proposed in [4], were designed to overcome the issue mentioned in the above paragraph.

2.4 Related Work

h_{t-1}

Chapter 3

Method

3.1   Dataset

3.2   Implementation

3.2.1   Conditional Random Fields

3.2.2   BI-LSTM Network

3.2.3   BI-LSTM-CRF Network

Chapter 4

Evaluation & Comparison

4.1   Results

Chapter 5

Conclusion

Bibliography

[1] CoNLL - State of the art. https://aclweb.org/aclwiki/CONLL-2003_(State_of_the_art). Accessed: 2019-12-11.

[2] Alan Akbik, Duncan Blythe, and Roland Vollgraf. Contextual string embeddings for sequence labeling. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1638–1649, 2018.

[3] James Hammerton. Named entity recognition with long short-term memory. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, pages 172–175, 2003.

[4] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[5] Zhiheng Huang, Wei Xu, and Kai Yu. Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*, 2015.

[6] John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.

[7] Zachary C Lipton, John Berkowitz, and Charles Elkan. A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*, 2015.

[8] Charles Sutton and Andrew McCallum. An introduction to conditional random fields. *Foundations and Trends® in Machine Learning*, 4(4):267–373, 2012.

[9] Erik F. Tjong Kim Sang and Fien De Meulder. Introduction to the conll-2003 shared task: Language-independent named entity recognition. In Walter Daelemans and Miles Osborne, editors, *Proceedings of CoNLL-2003*, pages 142–147. Edmonton, Canada, 2003.