

COMPARISON BETWEEN CONDITIONAL RANDOM FIELDS
AND LSTM NEURAL NETWORK APPROACHES FOR NAMED
ENTITY RECOGNITION

JOSEF GUGGLBERGER

SUPERVISED BY DIPL.-ING.CLEMENS SAUERWEIN, PhD

MASTERSEMINAR

LEOPOLD-FRANZENS-UNIVERSITÄT INNSBRUCK

DECEMBER 2019

Abstract

Contents

Abstract	iii
1 Introduction	1
2 Background & Related Work	3
2.1 Background	3
2.1.1 Named Entity Recognition	3
2.1.2 Conditional Random Fields	3
2.1.3 LSTM Neural Networks	4
2.2 Related Work	4
3 Method	5
3.1 Dataset	5
3.2 Implementation	5
3.2.1 Conditional Random Fields	5
3.2.2 BI-LSTM Network	5
3.2.3 BI-LSTM-CRF Network	5
4 Evaluation & Comparison	7
4.1 Results	7
5 Conclusion	9
Bibliography	11

Chapter 1

Introduction

Chapter 2

Background & Related Work

2.1 Background

2.1.1 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 [3]. For example, the terms *U.S.A* and *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, as for example in [4].

According to AclWeb [1], the current state of the art approach on the CoNLL-2003 [4] dataset reaches a F1 score of 93.09 %.

2.1.2 Conditional Random Fields

A *Conditional random field*, initially proposed in [2], 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 a 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 way that express some characteristic that is present in the training data and we want our model to hold. A feature function has to form:

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

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 [3], is then:

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

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

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

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.

2.1.3 LSTM Neural Networks

2.2 Related Work

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_\(State_of_the_art\)](https://aclweb.org/aclwiki/CONLL-2003_(State_of_the_art)). Accessed: 2019-12-11.
- [2] John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.
- [3] Charles Sutton and Andrew McCallum. An introduction to conditional random fields. *Foundations and Trends® in Machine Learning*, 4(4):267–373, 2012.
- [4] 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.