Named Entity Recognition and Classification using Artificial Neural Network

Luka Bašek
Fakulteta za elektrotehniko, računalništvo in informatiko
Koroška cesta 46
2000 Maribor, Slovenija
luka.basek@student.um.si

Borko Bošković
Fakulteta za elektrotehniko, računalništvo in informatiko
Koroška cesta 46
2000 Maribor, Slovenija
borko.boskovic@um.si

ABSTRACT

In this paper, we will analyze variants of Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) based models for sequence tagging, special for Name Entity Recognition (NER). The models which will be analyzed include LSTM network, bidirectional LSTM network (BI-LSTM), GRU network and bidirectional GRU network (BI-GRU) and using pre-trained GloVe vectors [1], Part of speech and characters embedding for features. We evaluated our models with data set for sequence labeling task CoNLL 2003 corpus. We obtain F1 score on this dataset - 86.04% [2].

Keywords

Natural Language Processing, Named Entity Recognition, Neural Network, Recurrent Neural Network, LSTM, GRU, Keras, GloVe vectors

1. INTRODUCTION

By developing information technologies, people today have a quick and wide access to a large amount of data. Everyday we meet a lot of sources of knowledge such as social networks, news, reviews, blogs, etc. There is a lot of data but the data is simple and raw isolated facts which have some meaning [15]. In the world is more and more data every day and data have to be analyzed to become a useful information. The computer can quickly process data and store the information. However, the computers without human presence can't analyze text and provide information about text. For this purpose, computer science has developed a lot of methods and techniques for analyzing and discovering knowledge. The field of Natural Language Processing (NLP) is an interdisciplinary area at the crossroads of computer science, artificial intelligence and linguistics. Our goal is Information Retrieval, more specifically the task of Named entity recognition.

To solve the problem of Named entity recognition, a variety of machine learning algorithms were used. The set of algorithms was composed of several statistical models such as the Hidden Markov Models (HMM) Florian et al. (2003) [4], Maximal Entropy Models (MaxEnt) Chieu et al. (2003) [1], Conditional Random Fields (CRF) McCallum et al. (2003) [12]. There are some machine learning methods but in this paper, we decide to use deep learning methods. On CoNLL 2003 one of the solutions contain Recurrent Neural Networks (RNN) with LSTM cells Hammerton (2013) [7] and this is the start point of our work.

The task is to explore deep learning methods in NLP and more specific for Named Entity Recognition (NER). Back in the few years, the neural networks have proved to be effective in the NLP area such as Text Classification Zhang et al. (2015) [21], Sentiment Analysis Severyn et al. (2015) [17], Part of speech tagging Wang et al. (2015) [19].

In the case of solving the problem of Name Entity Recognition, Hammerton (2003) [7] using RNN with LSTM cells presented by Hochreiter et al. (1997) [8]. Huang et al. (2015) [9] presents more complex models for NER based on the LSTM network, Bi-directional LSTM network, and various combinations with the CRF layer.

In this paper, we propose a variety of recurrent neural networks based models which include LSTM networks, bidirectional LSTM networks (BI-LSTM), GRU networks and bidirectional (BI-GRU) networks. Beside network architectures, we include additional features to help neural networks learn about the context of words. We include pre-trained GloVe for word representation, Part of Speech tags and Characters Embedding. We evaluate our models on English data from CoNLL 2003 shared task Sang et al. (2003) [16]. Our best F1 score is 86.05%. Slightly worse than current attractive solutions which have F1 score 91.21%, Xuezhe et al. (2016) [11] but this is a good starting point for improvements.

2. MODELS

In this section, we describe the models used in this paper: LSTM, BI-LSTM, GRU, BI-GRU.

2.1 LSTM Network

For operating sequential data the RNN is the most used approach. RNN takes as input a sequence of vectors $\vec{x} = (x_1, ..., x_n)$ and return another sequence $\vec{h} = (h_1, ..., h_n)$ that represents some information about the sequence by one-time unit. Reading articles about RNN, in theory, it works

[1] https://nlp.stanford.edu/projects/glove/

[2] The code of the this project is available at: https://github.com/lbasek/named-entity-recognition

DOI: https://doi.org/10.26493/978-961-7055-26-9.65-70