Report

On

Name Entity Recognition

Prepared by

Vasu Hudka (18012011024)

Guided By

Prof. Ketan J. Sarvakar

B.Tech Semester VII

Nov-Dec -2021

Subject: Deep Learning

Subject code: 2CEIT78PE1

Department of Computer Engineering

U. V. Patel College of Engineering

Ganpat University, Ganpat Vidyanagar – 384012

Abstract:

It is often claimed that Named Entity recognition systems need extensive gazetteers--lists of names of people, organizations, locations, and other named entities. Indeed, the compilation of such gazetteers is sometimes mentioned as a bottleneck in the design of Named Entity recognition systems. I report on a Named Entity recognition system which combines rule-based grammars with statistical (maximum entropy) models. I report on the system's performance with gazetteers of different types and different sizes, it is sufficient to use relatively small gazetteers of well-known names, rather than large gazetteers of low-frequency names.

Problem Description:

Named Entity Recognition(NER) is a process of recognizing information units like names, including person, organization and location names, and numeric expressions including time, date, money and percent expressions from unstructured text. The goal is to develop practical and domain-independent techniques in order to detect named entities with high accuracy automatically. In this project, I will train a Deep Learning model for Named Entity Recognition using Python. I will use the Tensor flow and Keras library in Python.

Notebook used:

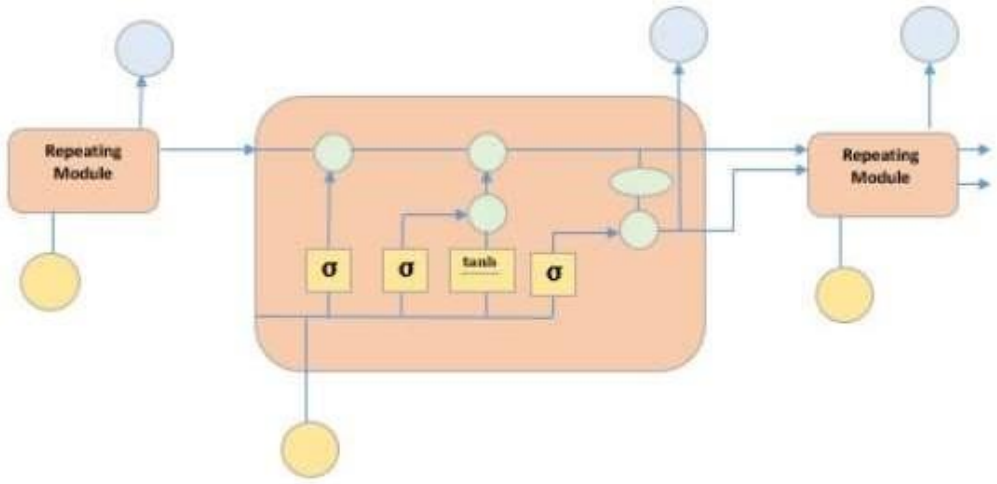
Google Colab - Colaboratory, or “Colab” for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education.

Technical requirements:

Sr no.	Name	Description
1	RNN	A recurrent neural network (RNN) is a type of artificial neural network which uses sequential data or time series data. These deep learning algorithms are commonly used for ordinal or temporal problems, such as language translation, natural language processing (nlp), speech recognition, and image captioning.
2	LSTM	Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can process not only single data points (such as images), but also entire sequences of data (such as speech or video).
3	TensorFlow	TensorFlow is a free and open-source software library for machine learning and artificial intelligence. It can be used across a range of tasks but has a particular focus on training and inference of deep neural networks.
4	Keras	Keras is an open-source software library that provides a Python interface for artificial neural networks. Keras acts as an interface for the TensorFlow library.

Architecture of LSTM:

It is special kind of recurrent neural network that is capable of learning long term dependencies in data. This is achieved because the recurring module of the model has a combination of four layers interacting with each other.



The picture above depicts four neural network layers in yellow boxes, point wise operators in green circles, input in yellow circles and cell state in blue circles.

An LSTM module has a cell state and three gates which provides them with the power to selectively learn, unlearn or retain information from each of the units.

The cell state in LSTM helps the information to flow through the units without being altered by allowing only a few linear interactions. Each unit has an input, output and a forget gate which can add or remove the information to the cell state.

The forget gate decides which information from the previous cell state should be forgotten for which it uses a sigmoid function.

The input gate controls the information flow to the current cell state using a point-wise multiplication operation of 'sigmoid' and 'tanh' respectively. Finally, the output gate decides which information should be passed on to the next hidden state

Work Flow:

Step 1: Import Required Libraries
Step 2: Read the data from csv file
Step 3: Pre-process the corpus
Step 4: Tokenizing the text into words
Step 5: Set length of the sequence to train
Step 6: Converting the texts into integer sequence
Step 7: Build lstm model
Step 8: Train the model
Step 9: Prediction

Dataset Name:

ner_dataset.csv

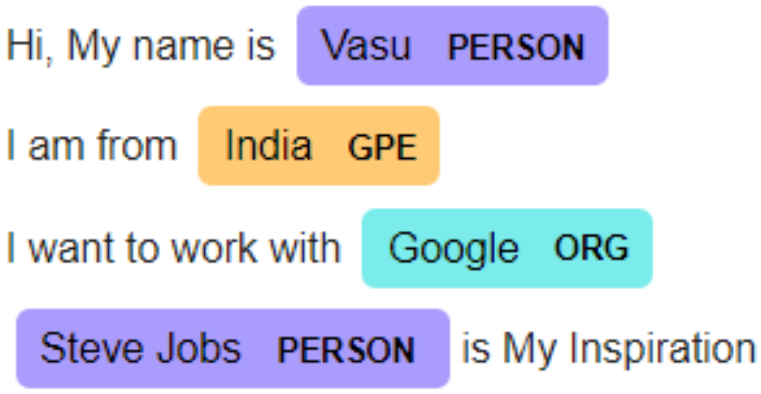
Dataset location:

https://drive.google.com/file/d/1YZY1cwz3pk31okCYfv5WERLXMQa2IhvH/view?usp=sharing

GitHub Code link:

https://github.com/vAsu27-git/Named-Entity-Recognition-NER-_DL

Output:


Hi, My name is Vasu PERSON
I am from India GPE
I want to work with Google ORG
Steve Jobs PERSON is My Inspiration