SyntaxNet: Neural Models of Syntax from Google
Technical Review

Yasaman Sabersheikh
2021-11-11

Technical Review: SyntaxNet: Neural Models of Syntax from Google

Overview

What is SyntaxNet?
SyntaxNet is a neural network framework in Natural Processing Language, that was released by Google is 2016. SyntaxNet tags words in a sentence with their syntactic part-of-speech and creates a parse tree showing dependencies between words in a sentence. SyntaxNet is a tensorflow based syntax parsing framework. SyntaxNet is free and for anyone to use.

How SyntaxNet Works

SyntaxNet takes a sentence as input, tags each word with a part-of-speech (POS) tag, it determines the syntactic relationships between words in the sentence, and represents the result as a dependency parse tree.

Per Google announcement in May 2016, SyntaxNet could be used to train any data, which come along Parsey McParseface, a trained English parser that can be used to analyze English text.[1]

What is Parsey McParseface? It is built on powerful machine learning algorithms that learn to analyze the linguistic structure of language, and that can explain the functional role of each word in a sentence.[1]

An input sentence is processed from left to right, with dependencies between words being incrementally added as each word in the sentence is considered. The model has over 94% accuracy, which makes it is really strong in NLP world. This accuracy is high and can be used in different applications.

Parsey McParseface was developed only for English when it was released. About

one year after, in 2017, SyntaxNet upgraded the code of Parsey McParseface to cover other languages and called it Parsey's Cousins. SyntaxNet could perform Text Segmentation and Morphological Analysis on 40 languages. [5]

Upgraded SyntaxNet

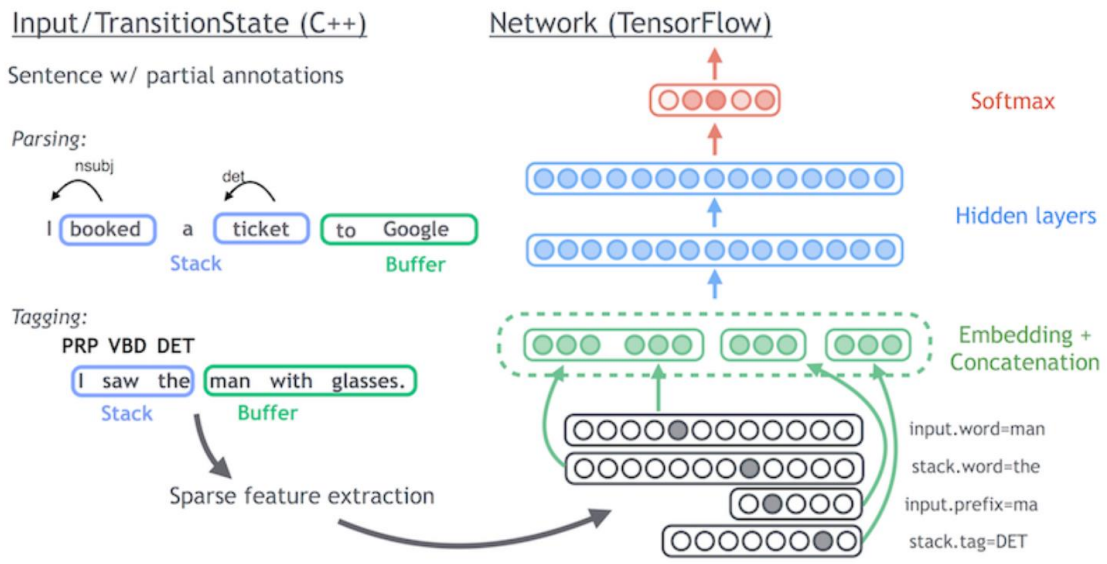
The upgrade to the original Parsey McParseface was a new technology that enables learning of layered representations of input sentences. The upgrade extends TensorFlow to allow joint modeling of multiple levels of linguistic structure, and to allow neural-network architectures to be created dynamically during processing of a sentence or document.[4]

The upgrade made it easy to build character-based models that learn to compose individual characters into words (e.g. 'c-a-t' spells 'cat'). By doing so, the models can learn that words can be related to each other because they share common parts (e.g. 'cats' is the plural of 'cat' and shares the same stem; 'wildcat' is a type of 'cat'). Parsey and Parsey's Cousins, on the other hand, operated over sequences of words. As a result, they were forced to memorize words seen during training and relied mostly on the context to determine the grammatical function of previously unseen words.[4] As the upgrade google released a set of new pretrained models called ParseySaurus. [4]

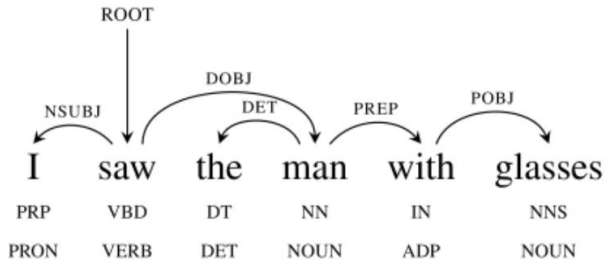
What SyntaxNet is Not [6]

SyntaxNet can not identify the sentiment of the sentence or does Sentiment Analysis. It can extract syntactic components from the sentences instead.
It cannot extract street address from a document (NER).
It is not an answer to all NLP problems.

An Example [8]



Dependency Parsing: Transition-Based Parsing



Setup

SyntaxNet Can be installed following the instruction provided in the following repository.
https://github.com/tensorflow/models/tree/f2f25096d3dc6561a855dab914cf2913100728d6/research/syntaxnet

Tutorial can be found in the following repository.
https://github.com/tiangolo/tensorflow-models/blob/master/research/syntaxnet/g3doc/syntaxnet-tutorial.md

Parsey McParseface API to create the parsed tree
https://deepai.org/machine-learning-model/parseymcparseface

Conclusion and Summary

The question we need to ask ourselves is why it is so hard for computers to parse a sentence meaningfully. The main reason would be the ambiguity and life experience that includes in human language.

SyntaxNet takes us one step closer to train computers to understand language like humans.

SyntaxNet is a framework for natural language syntactic parsers released by Google in 2016 and upgraded along the way.

Parsey McParseface is a SyntaxNet model trained on the English language. At its time of release, Parsey McParseface was the world's most accurate model of its kind.

Parsey's Cousins is a collection of pretrained syntactic models for 40 languages, capable of analyzing the native language of more than half of the world's population at often unprecedented accuracy

The upgrade to SyntaxNet Changed the parsing from word to character composing and then categorizing the words in the same family. These models use the character-based input representation mentioned above and are thus much better at predicting the meaning of new words based both on their spelling and how they are used in context.

References

1- SyntaxNet Google Announcement https://ai.googleblog.com/2016/05/announcing-syntaxnet-worlds-most.html
2- Google Paper on SyntaxNet original model https://arxiv.org/pdf/1603.06042.pdf
3- SyntaxNet Code and Setup https://github.com/tensorflow/models/tree/f2f25096d3dc6561a855dab914cf2913100728d6/research/syntaxnet
4- Google article on upgrade to Parsey McParseface and Parsey's Cousins https://ai.googleblog.com/2017/03/an-upgrade-to-syntaxnet-new-models-and.html
5- https://ai.googleblog.com/2016/08/meet-parseys-cousins-syntax-for-40.html
6- https://algorithmia.com/blog/advanced-grammar-and-natural-language-processing-with-syntaxnet
7- Parsey McParseface API - https://deepai.org/machine-learning-model/parseymcparseface
8- SyntaxNet Tutorial - https://github.com/tiangolo/tensorflow-models/blob/master/research/syntaxnet/g3doc/syntaxnet-tutorial.md