

# TensorFlow – NLP, an Overview

Prakash Manden

[pmanden2@illinois.edu](mailto:pmanden2@illinois.edu)

## What is TensorFlow?

TensorFlow is an end-to-end open source platform for machine learning<sup>1</sup>. It provides ML/AI developers a platform with various tools and libraries that makes it easy to manage all aspects of creating and deploying models. A powerful and flexible platform is provided that facilitates experimentation with building, evaluating and fine-tuning models. A model management pipeline that supports all steps in model generation and publishing (such as data ingestion, data validation/curation, transformation, training, evaluation and publishing) production is also provided (TensorFlow Extended (TFX)). It offers the flexibility to deploy models anywhere – on customer hosted servers, edge IOT devices or on web servers.

TensorFlow supports 2 modes of operations. 1) Graph based – In this mode a data structure ‘graph’ which represents Objects (or data) and Operations (units of computation) is created. Graphs can be saved, run and modified as needed. Graphs offer greater flexibility, eliminates dependencies on Python, (and hence can run on any device that doesn’t have a Python interpreter) and in most cases offer better performance through parallel execution. 2) Eager Execution – In this mode, code runs within a Python interpreter environment and executes code immediately. GPU and TPU can be used in this mode improving performance.

TensorFlow has a very strong community of researchers, developers and ML/AI engineers, that contribute, collaborate to develop and support this platform.

TensorFlow can be installed as a Python package or as a Docker Container. Google Colab provides a Jupyter Notebook environment that requires no configuration and can be run entirely on Google Cloud.

To summarize, TensorFlow is an increasingly popular and powerful platform that allows ML/AI engineers to build and deploy machine learning models that support a variety of use cases with a strong community to development support and an ecosystem of add on projects that continues to expand the capabilities to solve real world problems.

## What is NLP?

Natural Language Processing the study of processing data written or spoken in languages used by humans (hence natural) so that computers and algorithms can understand the meaning of data to solve various real world problems. For example, ‘sentiment analysis’ allows companies to understand how customer like or dislike their products by processing product reviews and other customer interactions. ‘Smart assistants’ allows computers to act as intelligent systems to respond to questions. Computers can be used to translate text from one language to another, etc.

Natural Language processing typically involves the following high level tasks – 1) Tokenization (i.e. identifying words from text) 2) Lexical Analysis (POS tagging) (i.e. categorizing words in syntactic

functions such as nouns, pronouns, verbs, adjectives, etc.) 3) Syntactic Analysis (Parsing) 4) Semantic Analysis 4) Pragmatic Analysis.

NLP is an actively researched and evolving area of research currently and is considered still in the early stages of development. As stated by Prof Cheng, “Robust and general NLP tends to be “shallow”, while “deep” understanding doesn’t scale up”, summarizes the state of the art of NLP technologies.

## NLP Capabilities in TensorFlow

TensorFlow is a powerful platform to develop NLP focused machine learning models. Keras APIs provide higher level classes and functions required to solve many of the natural language processing problems.

For example, as mentioned above one of the basic tasks of natural language processing is tokenization. Keras provides APIs for tokenization (*tensorflow.keras.preprocessing.text*) which is very easy to use.

Similarly, APIs are provided for Sequencing (i.e. to create indexes for a sequence of words – *text\_to\_sequences*), Padding (create indexes of symmetrical sizes - *tensorflow.keras.preprocessing.sequence*).

TensorFlow makes it very easy to build neural network models for text embedding (*tensorflow.keras.layers.Embedding*) and classification, which can also further be used for sentiment analysis (product reviews for e.g.). Creating a model for embedding can be done with a few lines of code as below (which is more or less the same for all types of models):

```
model = tf.keras.Sequential([
    tf.keras.layers.Embedding(vocab_size, embedding_dim, input_length=max_length),
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(6, activation='relu'),
    tf.keras.layers.Dense(1, activation='sigmoid')
])
```

TensorFlow naturally also provides APIs for training the models that are also equally easy to use!

TensorFlow also supports RNN, LSTM, Convolutional networks and other models that can be used for text sequencing/next word predictions. An example model creation for LSTM below:

```
model = tf.keras.Sequential([
    tf.keras.layers.Embedding(tokenizer.vocab_size, 64),
    tf.keras.layers.Bidirectional(tf.keras.layers.LSTM(64)),
    tf.keras.layers.Dense(64, activation='relu'),
    tf.keras.layers.Dense(1, activation='sigmoid')
])
```

‘TensorFlow Text’ also provides a collection of Classes and APIs that can be helpful for preprocessing (tokenizer, n-gram word splits, etc.) data for text based models.

TensorFlow can be used to build more sophisticated text based models (such as BERT) for sentiment analysis, named entity extractions, and intelligent chat bots.

## TensorFlow Data Services

Interestingly, TensorFlow also offers a [data service](#), which provides access to public research datasets that are useful for experimentation. Public datasets can be directly accessed via APIs (*tf.data.Dataset*). Sample datasets relevant to text processing includes - `ag_news_subset`, `anli`, `blimp`, `bool_q`, `c4`, `cfq`, `civil_comments`, `clinc_oos`, `cos_e`, `definite_pronoun_resolution`, etc.

In summary, TensorFlow is a very powerful machine learning platform that can be used to build models to solve real world NLP problems in an easy and effective manner.

## References

1. TensorFlow Website (<https://www.tensorflow.org/>)
2. TensorFlow data services (<https://www.tensorflow.org/datasets>)