

Deep Dishes – An AI Recipe Generator

“It was the secrets of heaven and earth that I desired to learn; and whether it was the outward substance of things or the inner spirit of nature and the mysterious soul of man that occupied me, still my inquiries were directed to the metaphysical, or in its highest sense, the physical secrets of the world.”

Mary Shelley, Frankenstein, Chapter 2

“Do not train a child to learn by force or harshness; but direct them to it by what amuses their minds, so that you may be better able to discover with accuracy the peculiar bent of the genius of each.”

Plato, The Republic, Book VII

Natural Language Processing

Natural Language Processing (NLP) requires that textual data should be rendered as digital or binary input. NLP uses the binary input, attaches weights to words, eliminates variations and redundancies, and then groups elements of a text, so that the algorithmic model may “read” the text and the relationships between its different elements. Such processes represent the *extraction* of meanings and relationships from textual data.

Conversion of textual data to usable data structures is important. Python, with modules, such as SciKit Learn and NLTK (Natural Language Toolkit), is capable of the processes of extraction. A textual document can become the data structure of a string. A *corpus* is then a collection of multiple documents, or a list of strings. With modules, such as PyTorch and TensorFlow, Python is also able to use neural networks in the processes of extraction, with the *tensor* as the necessary data structure. Here are some processes of extraction, more generally relevant to modules, such as SciKit Learn and NLTK:

Extraction

- **Vectorization – Tokenization** – Conversion of textual data to binary values in vectors of a matrix (list of lists), or in tensors of a neural network
- **TF – IDF** – Term Frequency – Inverse Document Frequency – Frequency of term divided by the number of documents with the term – Reduces the weight of terms with high frequency – Neural networks will use attention mechanisms to assign weights to terms
- **Stemming – Lemmatization** – Terms with inflections are converted to a common stem or lemma, and will not be processed as separate terms – Assists with the correct weight of terms
- **Stopwords** – Removal of common words, such as “a” and “the” – Assists with weighting terms
- **POS** – Assignment of parts of speech – Assists with lemmatization and the creation of a syntax for the document
- **Chunking** – Groups terms into meaningful units

With the completion of extraction, a model now trained on a text is capable of various processes:

Applications of an NLP Model

- **Classification** – Analysis of text in order to assign tags or categories to text based on its content
- **Sentiment Analysis** – Identification and extraction of opinions from text
- **Text Prediction and Text Generation** – Prediction and then generation of the word or words with the highest statistical probability of following the initial series of words

What is GPT-2?

GPT-2 or the Generative Pretrained Transformer 2 is an algorithm whose express goal is to understand the intricacies of the English language. The creators sought to read text through Natural Language Processing (NLP) in order to perform reading comprehension, text completion, summarization, and question answering with little supervision.

OpenAI offered a partial release of GPT-2 in March 2019, and complete release in November 2019. Although described in promotions as the ImageNext moment of Natural Language Processing (ImageNet is a database of 14 million images that revolutionized visual object recognition software), GPT-2 builds upon ULMFit, a transfer-learning process that attempts to extend supervised learning to the solution of classification problems without enough labelled data. Analogous to BERT from Google, which employs a bidirectional model, GPT-2 maintained the unidirectional model of GPT, but employed more GPUs in comparison to BERT in order to achieve greater computational power.

Aspects of GPT-2

Generative – Predicts and generates the next token in a sequence in an unsupervised manner. Given raw text, the transformer uses the statistics of features to predict and create more text

Pretrained – A large and powerful model trained with 1.5 billion parameters on a text corpus of 40 gigabytes, for several month on 100+ GPUs. TensorFlow allows for the processes of extraction

Transformer – A neural network architecture that utilizes dictionaries in order to eliminate iteration through sequences. With the efficiency of dictionaries, the time eliminated in iteration allows for the addition of more layers to the model.

Bigger is Better

Since GPT-2 offers greater computation, researchers do not need to insert too much knowledge into their models, and GPT-2 is then able to be more general in order to perform more tasks. Richard S. Sutton, in his essay, “The Bitter Lesson” (<http://www.incompleteideas.net/IncIdeas/BitterLesson.html>), notes that while the insertion of knowledge into models initially provides significant gains, such addition inevitably reaches a plateau that is only able to be overcome by means of scaling computation. GPT-2

provides evidence of the importance of the scale of networks and processing to the solution of problems. While GPT-2 reaches plateaus in summarization, GPT-2 avoids such plateaus in reading comprehension, question answering, and translation.

Within the context of NLP generally, and NLP with neural networks specifically, understanding of a text equates to the ability to predict the next word well:

Understanding = Prediction

Only when a model is capable of statistical calculations that approximate knowledge of vocabulary and syntax will predictions approach an accuracy that evokes understanding. With an accurate prediction, the generation of text is merely the feeding of the prediction to the model to release as output.

Within the context of neural networks specifically, attention assumes an essential role. Attention is a mechanism of the architecture of a neural network that manages and quantifies the interdependence of terms.

General Attention = Interdependence of Input and Output

Self-Attention = Interdependence within the Input

Attention is essentially a neural dictionary that performs queries for keys and returns values. Within a neural network, attention is a dictionary of dictionaries that references previous input, and since keys of the dictionary can be referenced without the need for iteration through data structure of sequences, such as strings, lists, arrays, tuples, and ranges, the dictionary allows for more expedient access and weighting of terms.

Our Objective

Our project endeavors to perform unsupervised learning with the neural network of the GPT-2 pretrained model. With an engineered dataset of recipes, selected for features of ingredients, name, description, and steps, our group trained the model in a Google Collaboratory Notebook that utilized TensorFlow and the gpt2-simple library. Our training consisted in 20,000 steps, and ultimately resulted in 1,000 generated recipe samples.