Amazon Comprehend – A Technical Review

Joel Kopp (joelk2) – November 12, 2020

Introduction

At the 2017 AWS re:Invent conference, Amazon introduced[1] their natural language processing service Amazon Comprehend as part of their machine learning suite of tools. It is billed as an NLP "service that uses machine learning to find insights and relationships in text" where "[n]o machine learning experience [is] required".[2] Indeed, the service utilizes a continuously updated pre-trained model[3] to gather insights on any document or set of documents.

In this review, we will examine the functionality of Amazon Comprehend in-depth, via the graphical user interface and the API. We will also use Amazon Comprehend with its built-in models to evaluate the sentiment on a test set of movie reviews from IMDB and compare its accuracy with a TensorFlow NLP model specifically trained on a segregated training set of IMDB reviews.

Cost[4]

Charges are based on a unit of text, defined as 100 characters. For their built-in models, Amazon's 12-month free tier allows analysis of up to 50,000 units of text, or 5 million characters, before charges are incurred. For sentiment analysis, once surpassing their free-tier allowance, clients are charged $0.0001 per unit for anything under 10 million monthly units. Bulk savings can then be realized as pricing is halved after the 10-million-unit limit and further at the 50-million-unit limit.

Customization is provided through user-defined models relayed through managed endpoints. Model training costs $3 per hour, and $0.50 per month for model management. For asynchronous classification using custom models, Amazon charges $0.0005 per text unit. For synchronous classification, Amazon charges $0.0005 per inference unit per second.

Features

Out of the box, Amazon Comprehend is quite easy to use in its graphical user interface form. Once an account has been set up, the user launches Amazon Comprehend from their AWS console. Upon launch, you are presented with a simple input text form, as seen in *Figure 1*. Here you select the built-in model, or the endpoint to your custom model, enter the text you wish to gain insight on, and click "Analyze".

---

[1] Amazon Web Services. (2018, January 12). *AWS re:Invent 2017 – Introducing Amazon Comprehend* [Video]. YouTube. https://www.youtube.com/watch?v=hdXvVyVjPLg

[2] AWS. (2020, November 12). Amazon Comprehend. Retrieved from https://aws.amazon.com/comprehend/

[3] Amazon Comprehend Developer Guide. (2020, November 12). How It Works. Retrieved from https://docs.aws.amazon.com/comprehend/latest/dg/how-it-works.html

[4] AWS. (2020, November 12). Amazon Comprehend Pricing. Retrieved from https://aws.amazon.com/comprehend/pricing/

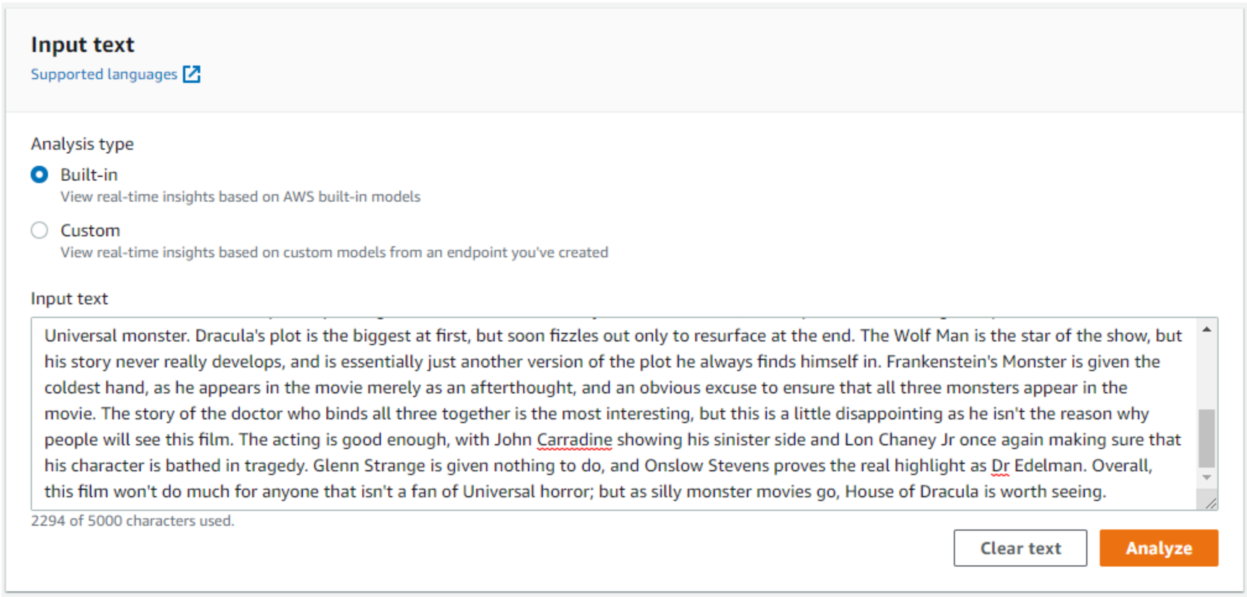

Figure 1: Input Text Form

From here, Amazon Comprehend will compute six NLP insights, including entity tagging, key phrase identification, language determination, personally identifiable information capture, sentiment analysis, and syntax/POS classification.

Sentiment

For sentiment classification, Amazon Comprehend will categorize the sentiment of your text with confidence intervals into one of four classifications; neutral, positive, negative, and mixed. *Figure 2* shows the sentiment analysis results for a sample IMDB movie review.

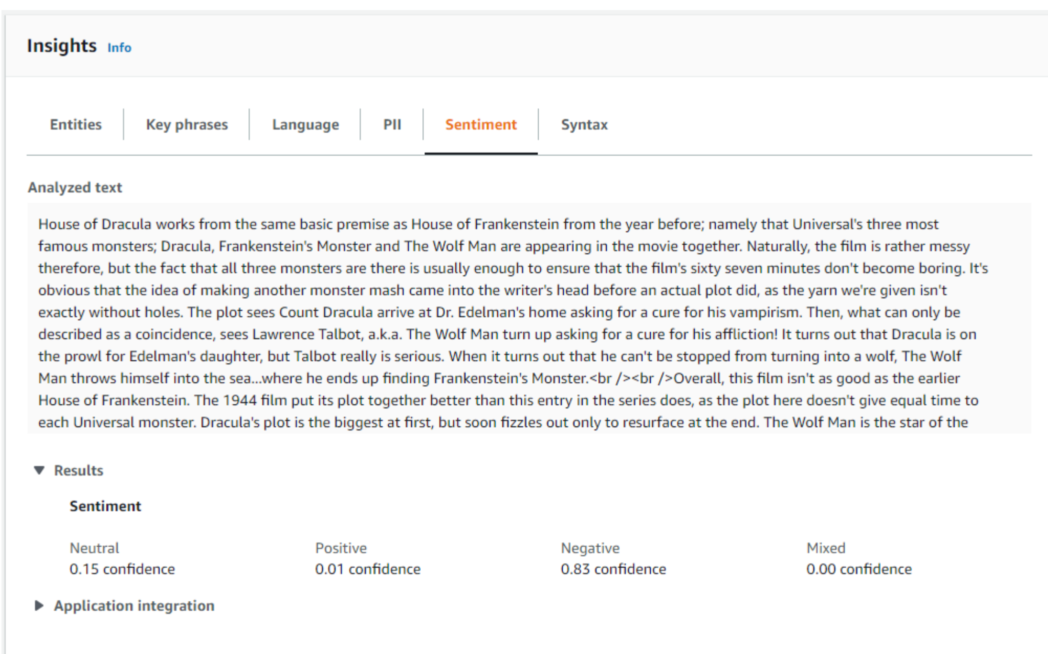

Figure 2: Results of Sentiment Analysis

The results indicate with 83% confidence that this is a negative review, which lines up with the true sentiment. Amazon Comprehend gives a 15% chance that this is a neutral review.

PII

Another useful feature is personally identifiable information recognition, or PII, which allows you to extract names, dates, social security numbers, bank routing numbers, or other bits of information that may be used to identify someone or compromise personal data. *Figure 3* shows the results of this type of analysis with confidence intervals.

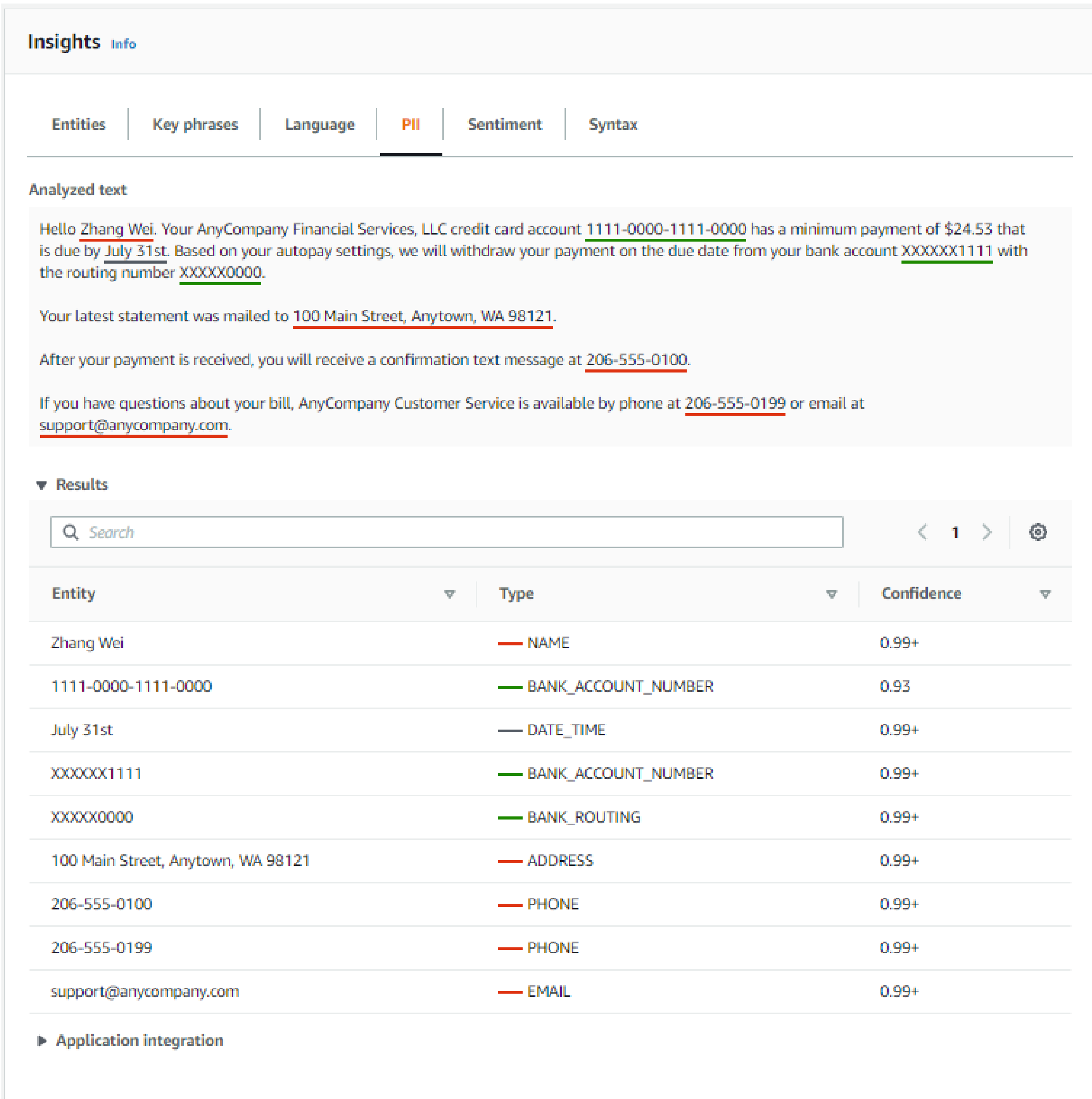

Figure 3: Results of PII Analysis

Syntax/POS Tagging

Amazon Comprehend also has robust syntax tagging. The service will label the grammatical parts of speech for each word and phrase in the document, with confidence intervals. *Figure 4* details the results of this functionality.

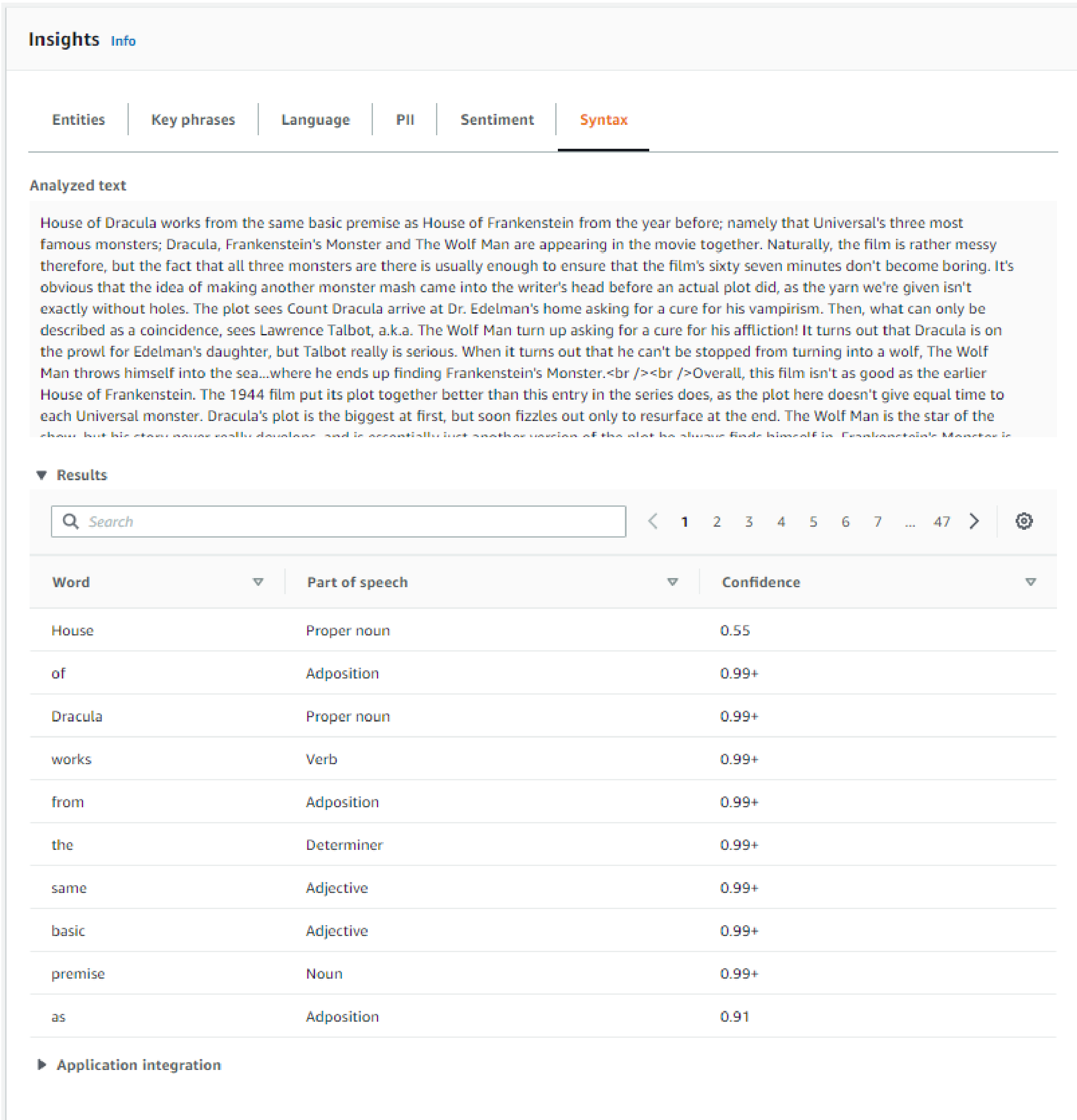

Figure 4: Results of Syntax/POS Tagging

Entity/Key Phrase/Language Recognition

Amazon Comprehend allows you to identify specific entities like a person, organization, quantity, date, or location through its entity recognition functionality. Further, Amazon Comprehend can identify the language of your document with confidence intervals. Amazon Comprehend also performs some rudimentary phrase extraction, as seen in Figure 5.

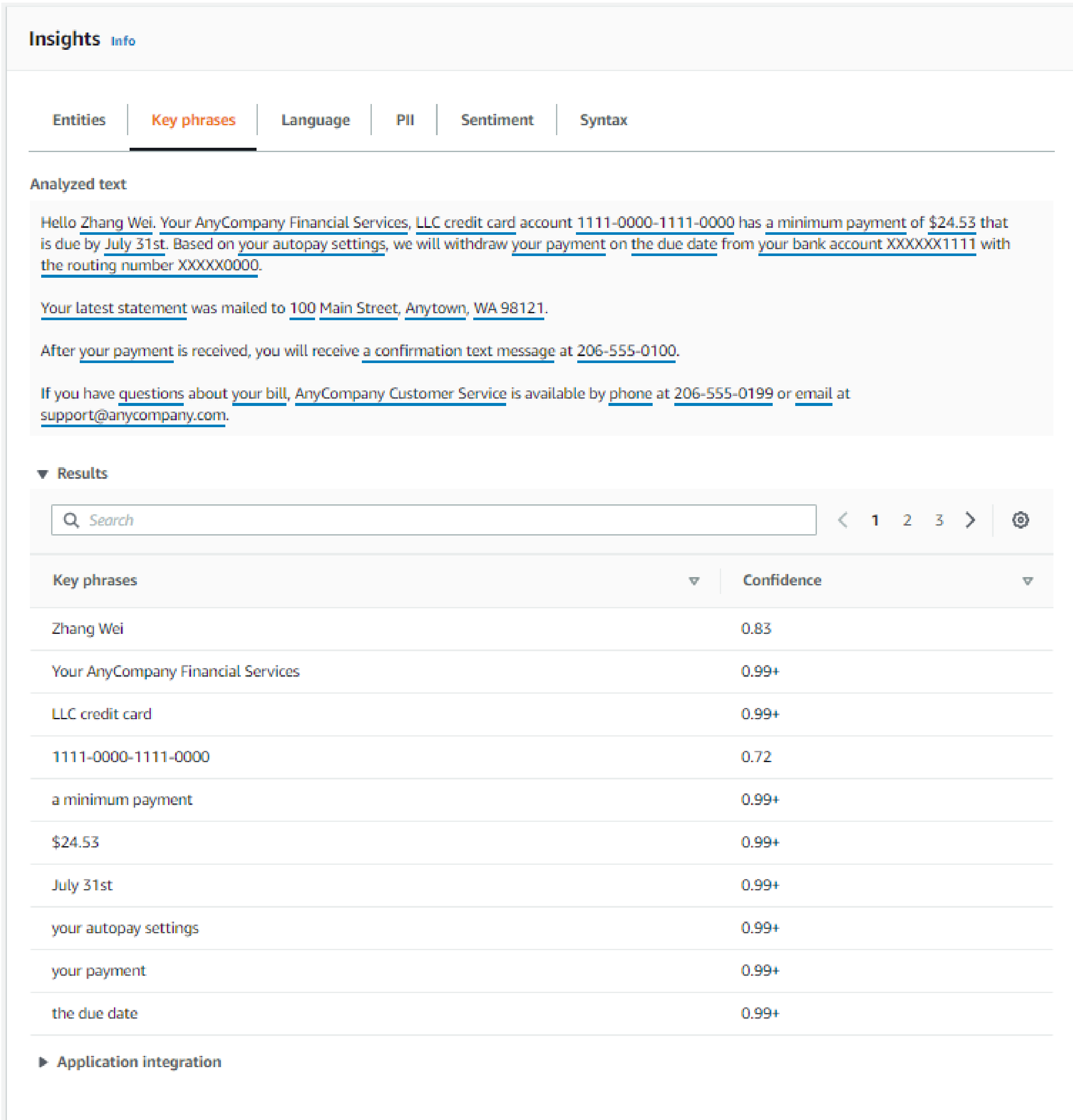

Figure 5: Results of Key Phrase Extraction

Limitations

Aside from the cost of the service, Amazon Comprehend only allows documents of at maximum 5,000 characters in length.

Accuracy

To evaluate the performance of Amazon Comprehend and its built-in models, we will evaluate the sentiment of a test set of 100 IMDB movie reviews through the AWS API and compare its accuracy to a TensorFlow model trained on 25,000 IMDB movie reviews. This test set is separate and not part of the 25,000-review training set, and can be found with their corresponding labels here: https://github.com/joel515/tech_review/blob/master/truth.csv.

TensorFlow Model

The TensorFlow model is built directly from TensorFlow's text classification tutorial.[5] The language model is pulled from the TensorFlow Hub here: https://tfhub.dev/google/tf2-preview/gnews-swivel-20dim/1. The model is composed of two hidden layers and is optimized using binary cross-entropy loss function. The model is trained using the 25,000 IMDB movie reviews, 10,000 of which being used as a validation set. The fitted model in the tutorial is evaluated using a 25,000 review test set. The 100 reviews used for comparison purposes is pulled from this test set.

The model is trained using batches of 512 samples for 20 epochs. The training concludes when we reach a minimum loss to avoid overfitting. At this point, the training set accuracy is 93%, while the test set accuracy is 86%. Finally, we evaluate the 100 segregated reviews used for comparison and obtain an accuracy of 82%. A Jupyter notebook with evaluations can be found here: https://github.com/joel515/tech_review/blob/master/tf_sentiment.ipynb.

Amazon Comprehend API

The Python Boto3 library was installed to use the AWS API functionality. The client is set up and the sentiment of review can be obtained using the following Python code:

```
client = boto3.client("comprehend")
response = client.detect_sentiment(Text=review, LanguageCode="en")
```

Here, `review` is a variable containing the text review string, and the `LanguageCode` is set to English (`"en"`). The `response` variable holds a Python dictionary containing a text string identifying the overall sentiment ("POSITIVE", "NEGATIVE", "NEUTRAL", and "MIXED"), as well as the confidence intervals for each. To compare to our binary positive/negative categorization from TensorFlow, we classify a neutral or mixed review as positive if the positive confidence interval is greater than the negative confidence interval, and vice-versa. The Python script used to evaluate the reviews through the API can be found here: https://github.com/joel515/tech_review/blob/master/aws_sentiment.py.

Looping through the 100 IMDB comparison test reviews and comparing them to truth, we find the Amazon Comprehend accuracy is 73%. The evaluations can be found here: https://github.com/joel515/tech_review/blob/master/aws_response.csv.

Conclusion

While its sentiment accuracy is not as good as a custom-trained model (73% to TensorFlow's 82%), the convenience of an adequate built-in model is hard to beat. Limiting factors include its cost

[5] TensorFlow Hub Tutorials. (2020, November 12). Text Classification with Movie Reviews. Retrieved from https://www.tensorflow.org/hub/tutorials/tf2_text_classification

(after the free tier) and a 5,000 character document limit. This last limitation was encountered in this small evaluation as one of the 100 IMDB reviews had to be replaced with a smaller one as it was over this character limit. However, with Amazon Comprehend, you do get five additional insights with the same built-in models, including POS tagging, language identification, key phrase parsing, entity recognition, and personally identifying information recognition. And finally, if you did want to build your own model, you have the enormous AWS infrastructure at your disposal.