Azure Cognitive Services – Text Analytics

Ed Pureza (epureza2), November 8, 2020, CS 410 Technology Review

Data science-based jobs have grown by as much as 15 to 20 times just in the past four years (Knowledge@Wharton, UPenn 2019). This massive growth has led to a high demand for data scientists with projected shortages of up to 250,000 beyond 2020 (DuBois 2019). As companies rush to get AI infused in their processes and products, many are finding it difficult to realize a return on investment with only 13% of data science projects making it into production (Venture Beat 2019). The shortage of data scientists along with a low implementation rate puts AI adoption at great risk. The market recognizes both issues of constrained resource supply and challenges with productionization, and vendors are attempting to provide solutions that make data science capabilities more accessible to everyone. One of the providers in this space is Microsoft and one of their offerings is Azure Cognitive Services. The product promises to "bring AI within reach of every developer without requiring machine-learning expertise." Along with vision, search, speech, and decision capabilities, they also offer language capabilities that aim to address needs for extracting meaning from unstructured data. One of the more interesting features in this category is text analytics. Azure text analytics can detect sentiment, key aspects, and opinions.

Getting started with Azure Text Analytics is simple and experimenting with it is affordable given a one-year free subscription to Azure. After creating an account and instantiating a text analytics service, the next step is the obtain an API key that can be used to access the Azure's text analytics service via API. Users can also experiment with the service using a web UI. The API request requires the language, an identifier, and the text. The API returns the overall sentiment with confidence scores for positive, neutral, and negative classification. It also breaks up the text by sentence and provides a sentiment score for each sentence in a similar way – with confidence scores for the three classifications. For each sentence, it identifies the aspects or main subjects and gives each aspect a sentiment score and a relationship type, which can be documents, sentences, or opinions. For an aspect with a related opinion, a listing for the word representing the opinion is provided along with the sentiment scores. In addition to sentiment analysis, Azure text analytics also identifies key phrases and provides named entity recognition which can categorize different entities in text into persons, locations, events, products, or organizations.

To achieve feature parity from scratch would require obtaining data in various languages, classifying the data, cleaning and structuring the data, model training and hyperparameter optimization, deploying the models, defining the API message structures, creating an endpoint for receiving text analysis requests and implementing load balancing to handle requests. To maintain this type of solution would require a product management team to manage the product life cycle and a development team to address continuous improvements to the service pipeline. Azure Cognitive Services have reduced the data science work involved in text analytics to an integration exercise – sending and receiving messages to the text analytics service.

Another alternative would be to use packages like TextBlob and MeTA for Python (and other languages). These packages also provide part-of-speech tagging, classification, sentiment analysis, and key phrase

extraction along with other text analytics capabilities. This option eliminates the need to have an Azure subscription, but the same considerations above persist. Development and maintenance costs would be substantial.

For options including Azure Cognitive Services, the work to procure the data to analyze remain the same. However, Azure Cognitive Services alleviates the work of gathering and maintaining training datasets and tuning of the models. The work left over is putting the text analytics service into production to return results for each request. Text analytics is now accessible to software engineers, reducing the need for heavy data science work and data scientists.

This area of text analytics has undoubtedly benefited from research in the field that has allowed development to focus more on development. The work of the relatively small number of experts with PhDs is now able to find its way to the masses with these tools. As the ecosystem around AI matures, the level of abstraction increases, and state-of-the-art capabilities have effectively become commoditized.

The task is now about putting solutions together and answering business questions about how sentiment analysis and opinion mining can be used. For example, product managers can hypothesize about the relationship between sales levels for new products and identifying the aspects that customers liked and disliked. This information could then be used for demand planning by adjusting projected sales based on initial customer reviews or responses. The sentiment analysis and topic mining can also be used by new product development when considering new features.

If Azure Cognitive Services delivers on its promises of reducing or eliminating the need for data science expertise to benefit from advances in AI, then the vision of incorporating data science into more products and processes will have a more viable path to success. These types of services will allow for faster experimentations and iterations of new products. Creativity in putting together these AI-based capabilities will be the main driver for benefit realization and machine learning adoption in the enterprise. Advances in this area will also allow data scientists to concentrate on harder problems to solve and continue to drive innovation. With sustained progress, hopefully the risk of disenchantment with AI can be reduce and “AI winters” can be avoided.

References

DuBois, Jen. 2019. *Is there a Data Scientist Shortage in 2019?* June 3. https://quanthub.com/is-there-a-data-scientist-shortage-in-2019/.

Knowledge@Wharton, UPenn. 2019. *What's Driving Demand for Data Scientists?* March 8. https://knowledge.wharton.upenn.edu/article/whats-driving-demand-data-scientist/.

Venture Beat. 2019. *Why Do 87% of data science projects never make it to production?* July 19. https://venturebeat.com/2019/07/19/why-do-87-of-data-science-projects-never-make-it-into-production/.