Technology Review – IBM Watson Natural Language Classifier

Overview:

IBM Watson Natural Language Classifier (NLC) enables easy development of text classification models without the need for having deep technology expertise in Data Science or Machine learning. It uses machine learning algorithms to return the top matching classes for short text inputs. The classifier can be trained with sample data and it can then make predictions about texts. NLC supports around 10 languages including languages such as English, Spanish and French.

Steps to create and access the service:

IBM Watson Natural Language Classifier (NLC) enables applications to analyze and understand texts and classify them. Following are the steps involved in creating and accessing the classifier.

1) Prepare training data – In this step, we need to prepare some sample data for training the classifier.
2) Create the classifier – We can use the APIs provided by the NLC service to create the classifier.
3) Train the classifier – Using the training data from step1 or step5, we can train the classifier. NLC provides APIs for training the classifier.
4) Query the classifier – We can use the NLC APIs to fetch the matching classes for a given text.
5) Evaluate results – In this step, we need to evaluate the results provided by the API and update the training data based on the results.

Best Practices and Guidelines for training:

The performance of the classifier is heavily dependent on the data provided to it for training. So we need to carefully select the text inputs and the classes expected. Following are some of the guidelines for training the classifier to achieve best results.

- The length of the input must be fewer than 60 words.
- The number of classes shouldn't be huge. It should be within few hundreds to achieve optimum results.
- Each class should be matched with at least 5 to 10 records so that there is sufficient training on that class.

Use cases of NLC:

The natural language classifier can be used in several industries and applications to fulfill different use cases of Text Classification. Some of them are as below.

- Social media – Perform sentiment analysis on Twitter feeds and classify them into categories.
- Resume analysis – Analyze resumes and job applications to derive insights.
- E-commerce – Helps users by narrowing the product search results using product classification.
- Education and Government – Useful for educational and government organizations to classify texts and organize them.
- Banking – Perform analysis on Risks, Transactions and Investments and classify them appropriately.

Advantages of IBM Cloud Platform:

IBM Watson Natural Language Classifier (NLC) is offered as an IBM Cloud Service. This offers several key advantages as explained below.

- Costs – There are no infrastructure costs as NLC is offered as a cloud service. This helps businesses to avoid infrastructure efforts and focus on building applications.
- High availability – This is supported since the service is offered through multiple IBM Cloud locations and there is no single point of failure.
- Scalability – The service is highly scalable and new instances can be easily added.

Conclusion:

IBM Watson Natural Language Classifier (NLC) is a very simple to use tool for text classification. The product claims that the users can get started with natural language processing very easily. I think the product lives up to the expectations. I was able to sign up for a free trial account and explore the features of the service very quickly. Since it is cloud based, there are no additional costs and efforts related to maintaining the infrastructure required for the application. The product has extensive documentation and how-to-use guides on the APIs. It helps to reduce the development time required for building a text classifier and addresses common use cases for text classification. Overall, I feel that this is a very simple to use product for text classification.

References:

- https://www.ibm.com/cloud/watson-natural-language-classifier
- https://cloud.ibm.com/docs/natural-language-classifier
- https://cloud.ibm.com/apidocs/natural-language-classifier