

Review of Azure Cognitive Services

Nitishj2@illinois.edu

Introduction (What)

Azure Cognitive Services offers various AI and machine learning APIs and SDKs, which can be used in applications to make them more intelligent. The APIs are split up into multiple categories, such as vision, speech, language, knowledge, and search APIs.

Available Language/Text Related Services and APIs (Features)

Language Understanding (LUIS)

LUIS can extract meanings from text. It uses machine learning to train the API. Can be used to Build natural language understanding into apps, bots, and IoT devices.

Most common AI interactions occur through bots, the API provides ability to interact intelligently with the user. We can build conversational intelligence into our bots by using Azure Language Understanding Intelligent Service (LUIS).

LUIS applies custom machine-learning intelligence to conversational natural-language text. LUIS uses certain aspects of the text to predict the user's overall meaning and pull out relevant detailed information. The applications can use this information to interact with the user.

LUIS makes use of three key aspects for understanding language:

- **Utterances:** An utterance is input from the user that the app needs to interpret.
- **Intents:** An intent represents a task or action the user wants to do; (purpose or goal expressed in a user's utterance)
- **Entities:** An entity represents a word or phrase inside the utterance that we want to extract. Entities aren't only words or phrases, but also simply data. This data helps provide specific context for the utterance and aids the algorithm in more accurately identifying the intent. Not every utterance contains entities, though.

For the following sample utterance: "Book 2 tickets on a flight to New York for New Year's Eve."
The API identified entities as:

New York: We can classify this entity as Location.Destination.

New Year's Eve: We can classify this entity as Event.

The number 2: This number maps to a built-in entity. In LUIS, such an entity is known as a prebuilt entity, specifically a prebuilt number.

QnA Maker

This offers a web API that trains an AI model to respond to user's questions in a more natural, conversational way. It can be trained using FAQ, URLs/documents, and more.

Text Analytics API

This API provides natural language processing over raw text and includes the features mentioned here. One of the features is language detection for up to 120 languages. It returns a language code and a score indicating the strength of the analyzation. It also provides Key Phrase Extraction, where it extracts key phrases to identify the main points in sentences, and it offers sentiment analysis, which we can use to find out what people think of certain things by analyzing text.

Translator Text API

This API can be used to provide text-to-text language translation in more than 60 different languages.

Immersive Reader

Help readers of all abilities comprehend text using audio and visual cues.

Content Moderator

The content moderator offers automatic moderator capabilities such as detecting possible adult and racy content in videos/text. It offers review tooling where automatic moderation can be used in conjunction with human involvement as well.

Other Cognitive Services APIs (Features)

- Vision:
 - **Computer Vision API:** This API provides image processing and recognition. We can use this API for categorizing images, tagging images based on content, recognizing handwritten text, flagging adult content, cropping images, detecting human faces, and more.
 - **Custom Vision Service:** This API offers a tool in order to build custom image classifiers. This can be used to identify images, such as certain flowers or dogs for instance. This API needs to be trained by uploading images to it.
 - **Face API:** This API can detect human faces in an image. It can extract information from the images, such as pose, facial hair, glasses, gender, age, and head pose. We can use this API for face verification, face grouping, face identification, and finding similar faces. Emotion API is also a part of the Face API and can be used to detect the emotion of people in images.
 - **Video Indexer:** This API can extract insights from your videos using various artificial intelligence technologies. This API is used inside Azure Media Services as well and will be covered in more detail later in this chapter.
- Speech:
 - **Speech to Text/Text to Speech API:** This API provides speech-enabled features, such as voice command control and speech transcription and dictation. It offers speech to text, where human speech is converted into text, which can be used to control applications by speech and text to speech, where text is converted to audio streams, which can be played back in applications.
 - **Speaker Recognition API:** This API provides speaker verification, which can be used to identify a person using voice commands for authentication. It provides speaker identification, where it can recognize a person in a group of speakers by their voice.
 - **Speech Translation API:** This API offers a service that translates conversational speech from one language into the text of another language.

- Web Search:
 - **Bing Search APIs:** This API consists of multiple APIs and functions. It includes all Bing search APIs, such as the Bing web search API, Bing image search API, Bing video search API, and Bing news search API.
 - **Bing Autosuggest API:** This offers suggestions when we are typing the first few characters of a word. We can use it to populate a drop-down box under search boxes.
 - **Bing Custom Search API:** This offers tailored search experiences. We can tailor results based on interests. So, instead of letting users go through pages to search for relevant content, we can filter irrelevant content before providing it to your users.
 - **Bing Entity Search API:** This API provides search results that includes entities and places that can be used for tourist attractions, for instance.

Text Classification and Moderation (Experiments)

Content Moderator API provides machine-assisted content moderation for images, text, and videos. When we are using machine-assisted content moderation, we can either block, approve, or review the content based on the policies and thresholds. We can use machine assistance to augment human moderation of environments where text content is generated (eg: Chat rooms, Discussion boards, Chatbots, E-commerce catalogs, Documents etc.)

The response from the Text Moderation API includes the following information:

- A list of potentially unwanted words found in the text.
- What type of potentially unwanted words were found.
- Possible personally identifiable information (PII) found in the text.

Profanity Detection

When we pass text to the API, any potentially profane terms in the text are identified and returned in a JSON response. The profane item is returned as a Term in the JSON response, along with an index value showing where the term is in the supplied text.

We can also use custom term lists with this API. In that case, if a profane term is identified in the text, a ListId is also returned to identify the specific custom word that was identified.

A sample JSON response is shown here:

```
"Terms": [
{
  "Index": 118,
  "OriginalIndex": 118,
  "ListId": 0,
  "Term": "crap"
}
```

Classification

This feature of the API can place text into specific categories based on the following specifications:

- Category 1: Potential presence of language that might be considered sexually explicit or adult in certain situations.

- Category 2: Potential presence of language that might be considered sexually suggestive or mature in certain situations.
- Category 3: Potential presence of language that might be considered offensive in certain situations.

When the JSON response is returned, it provides a Boolean value for a recommended review of the text. If true, we should review the content manually to determine the potential for any issues. Each category is also returned with a score between 0 and 1 to indicate the predicted category for the evaluated text. The higher the score, the more likely it is that the category might apply. Here is a sample JSON response:

```
"Classification": {
  "ReviewRecommended": true,
  "Category1": {
    "Score": 1.5113095059859916E-06
  },
  "Category2": {
    "Score": 0.12747249007225037
  },
  "Category3": {
    "Score": 0.98799997568130493
  }
}
```

Personally-identifiable-information

This feature of the API helps detect if any values in the text might be considered PII. Key aspects that are detected include:

- Email addresses
- US mailing addresses
- IP addresses
- US phone numbers
- UK phone numbers
- Social Security numbers

If possible PII values are found, the JSON response includes relevant information about the text and the index location within the text.

A sample JSON response is shown here:

```
"PII": {
  "Email": [{
    "Detected": "abcdef@abcd.com",
    "SubType": "Regular",
    "Text": "abcdef@abcd.com",
    "Index": 32
  }],
  "IPA": [{
```

```

    "SubType": "IPV4",
    "Text": "255.255.255.255",
    "Index": 72
  },
  "Phone": [{
    "CountryCode": "US",
    "Text": "5557789887",
    "Index": 56
  }, {
    "CountryCode": "UK",
    "Text": "+44 123 456 7890",
    "Index": 208
  }],
  "Address": [{
    "Text": "1 Microsoft Way, Redmond, WA 98052",
    "Index": 89
  }],
  "SSN": [{
    "Text": "999-99-9999",
    "Index": 267
  }]
}

```

Sentiment Analysis using Text Analytics API (Experiments)

Using this API, we can identify language, discover sentiment, extract key phrases, and detect well-known entities from text.

In a nutshell, if we give a piece of text to the Analytics service, it returns a score between 0 and 1 denoting overall sentiment in the input text. Scores close to 1 indicate positive sentiment, while scores close to 0 indicate negative sentiment. We do not need to train the model before use, as the model is already trained. In a nutshell, this is how this service works:

1. Obtain a large dataset of text with sentiment scores.
2. Split text into words and apply stemming (converting word to its root form) e.g. fishing, fisher or fished is reduced to fish.
3. Create features from words. Some of the key features used are:
 - a. N-Grams: Generate all possible combinations of n consecutive words e.g. for “we are learning ML” and n=2, the sequence would be “we are”, “are learning”, “learning ML”.
 - b. Part-of-speech tagging: It is the process of identifying words belonging to a particular part of speech. A simplified form of this is identification of words as nouns, verbs, adjectives etc.
 - c. Word embedding: It is the process of mapping syntactically similar words close to each other e.g. car and bike are closer to each other than are car and office.
4. Once the features have been identified, the classifier is trained with the features.

To call the API, we structure a POST request, send it to the /sentiment endpoint, and receive a JSON response that tells us a sentiment score.

Alternatives

IBM Watson: It combines artificial intelligence (AI) and sophisticated analytical software for optimal performance as a "question answering" machine.

TensorFlow: TensorFlow is an open-source software library for numerical computation using data flow graphs. Nodes in the graph represent mathematical operations, while the graph edges represent the multidimensional data arrays (tensors) communicated between them. The flexible architecture allows to deploy computation to one or more CPUs or GPUs in a desktop, server, or mobile device with a single API.

Keras: Deep Learning library for Python. Convnets, recurrent neural networks, and more. Runs on TensorFlow or Theano. <https://keras.io/>

scikit-learn: scikit-learn is a Python module for machine learning built on top of SciPy and distributed under the 3-Clause BSD license.

PyTorch: PyTorch is not a Python binding into a monolithic C++ framework. It is built to be deeply integrated into Python. We can use it naturally like we would use numpy / scipy / scikit-learn etc.

In addition to this, Amazon Cloud also offers similar services like Cognitive services.

Conclusion

Cognitive Services is part of the AI offering of Azure. It offers many APIs that can be consumed as-is (and some APIs that need training) and can be used to easily create our own custom AI solutions.

Azure Cognitive services provides a very comprehensive portfolio of domain specific AI capabilities in the market. Most services do not require to train the model or machine learning knowledge itself.