CS 410: Technology Review - Survey of Cloud Services for Sentiment Analysis

Stanley Ho (smho2@illinois.edu)

1. Introduction

Sentiment analysis is for analyzing an opinionated piece of text to determine whether the underlying sentiment is positive, negative, neutral, or mixed. It uses machine learning and natural language processing to determine the result. There are many applications, including analyzing customer feedback, capturing brand sentiment in social media to better understand the targeted audiences, and analyzing market sentiment towards a stock for investors or traders to plan long or short positions, etc.

The top 3 public cloud providers, *Amazon Web Services (AWS)*, *Microsoft Azure*, and *Google Cloud*, all offer cloud services for sentiment analysis. This technical review is a survey of their *sentiment analysis* cloud services.

2. Sentiment Classification

The classification in *sentiment analysis* is defined in two ways:

1. *Polarity analysis* to categorize the opinion in the text as positive, negative, or neutral.
2. *Emotion analysis* to characterize the precise feeling of the opinion holder, e.g. happy, sad, angry, surprised, and disgusted, etc.

The granularity of the *sentiment analysis* is also important:

1. The *overall sentiment* of the text.
2. The *entity sentiment* (i.e. opinion holder or opinion target) in the text.

These are important to understand the *sentiment analysis* offerings from the cloud providers.

3. Amazon Comprehend

Amazon Comprehend is a natural language processing (NLP) service offered by AWS, and it uses machine learning to discover insights from text. *Amazon Comprehend* supports *sentiment*

analysis via its *Sentiment Analysis API*, and it supports *polarity analysis* and *emotion analysis* for the sentiment classification.

The *Sentiment Analysis API* takes a text as input, and returns the overall sentiment of the text (i.e. Positive, Negative, Neutral, Mixed). The text input must be under 5000 bytes of UTF-8 encoded characters, and the text must be in one of the twelve supported languages, including English (en), Spanish (es), French (fr), German (de), Italian (it), Portuguese (pt), Arabic (ar), Hindi (hi), Japanese (ja), Korean (ko), Chinese-Simplified (zh), and Chinese-Traditional (zh-tw).

The *Sentiment Analysis API* returns two pieces of information with respect to the *overall sentiment* of the text. It returns a list of sentiments each with a confidence score representing the sentiment detection accuracy, and the confidence scores range from 0 to 1. It also returns the inferred *overall sentiment* which is simply the sentiment with the highest confidence score. The sentiment detection accuracy varies between datasets. The returned sentiment detection is based on the entire text; there is no granular sentiment information returned for the entity (i.e. opinion holder or opinion target) within the text.

For programming, REST API and command line interface (CLI) are available. Client libraries for Java, Python, .NET, JavaScript, Go, Rudy, PHP are also supported.

For pricing, the requests are measured in units of 100 characters as follows:

Number of requests	Price Per Unit (1 Unit = 100 characters)
Up to 10M units	$0.0001
10M - 50M units	$0.00005
> 50M units	$0.000025

4. Microsoft Azure

Text Analytics is a text-mining AI service in *Microsoft Azure* that uncovers insights such as sentiment analysis, entities, relations, and key phrases in unstructured text. The service supports *Sentiment Analysis* based on *polarity analysis* and *emotion analysis* in two levels:

1. *Overall sentiment analysis* (aka *Sentiment Analysis* in *Text Analytics*) takes text input, and returns the overall sentiment (i.e. negative, neutral and positive) at a sentence and document level, with a confidence score for each. The confidence scores range from 0 to 1, and the confidence scores within each document or sentence add up to 1.
2. *Entity (or aspect-based) sentiment analysis* (aka *Opinion Mining* in *Text Analytics*) takes text input, and returns sentiment specific to the entity (i.e .opinion holder or target) identified in the text. For example, the text "The room was great, but the staff was unfriendly." would result in negative sentiment in overall sentiment analysis, but the *entity sentiment analysis* would separate the sentiment per entity.

3. The actual accuracy of the sentiment analysis varies between datasets.

The text input must be under 5,120 characters. For *overall sentiment analysis*, the text must be in one of the fifteen supported languages, including English (en), Spanish (es), French (fr), German (de), Italian (it), Portuguese Portugal (pt-PT), Portuguese Brazil (pt-BR), Turkish (tr), Dutch (nl), Norwegian (no), Hindi (hi), Japanese (ja), Korean (ko), Chinese-Simplified (zh-hans), and Chinese-Traditional (zh-hant). For *entity sentiment analysis*, only English (en) is supported.

For programming, REST API and command line interface (CLI) are available. Client libraries for C#, Java, Python, JavaScript, Go, and Ruby are also supported.

For pricing, the requests are measured in units of 1,000 characters as follows:

Number of requests	Price Per 1,000 Units (1 Unit = 1,000 characters)
Up to 5K units	Free
Up to 0.5M units	$1.00
0.5 - 2.5M units	$0.75
2.5M - 10M units	$0.30
> 10M units	$0.25

5. Google Cloud

Natural Language API is a service in *Google Cloud* which enables developers to easily apply natural language processing to their applications. It supports sentiment analysis based on *polarity analysis and emotion analysis* in two levels:

1. *Overall sentiment analysis* (aka *Sentiment* in *Natural Language API*) takes text input, and returns the overall sentiment (i.e. positive, negative, or neutral), with a score indicating the overall emotion of the text, and a magnitude indicating how much emotion content is present in the text.
2. *Entity sentiment analysis* (aka *Entity Sentiment* in *Natural Language API*) takes text input, and returns sentiment specific to the entity (proper nouns and common nouns) in the text. Entity sentiment is represented by a score and magnitude and is determined for each mention of an entity; these scores are then aggregated into an overall sentiment score and magnitude for an entity.
3. The actual accuracy of the sentiment analysis varies between datasets.

The text input must be under 1,000,000 bytes with at most 5,000 entities within the text. For *overall sentiment analysis*, the text must be in one of the sixteen languages, including English

(en), Spanish (es), French (fr), German (de), Italian (it), Portuguese Brazil and Continental (pt), Turkish (tr), Dutch (nl), Arabic (ar), Indonesian (id), Japanese (ja), Korean (ko), Thai (th), Vietnamese (vi), Chinese-Simplified (zh), and Chinese-Traditional (zh-Hant). For *entity sentiment analysis,* the text must be in one of the three languages, including English (en), Japanese (ja), or Spanish (es).

For programming*, REST API* and command line interface (CLI) are available. Client libraries for Java, Python, C#, Go, JavaScript, PHP, and Rudy are also supported.

For pricing, the requests are measured in units of 1,000 characters as follows:

Number of requests	Price Per 1,000 Units (1 Unit = 1,000 characters)
Up to 5K units	Free
5K - 1M units	$1.00 for overall sentiment analysis; $2.00 for entity sentiment analysis
1M - 5M units	$0.50 for overall sentiment analysis; $1.00 for entity sentiment analysis
5M - 20M units	$0.25 for overall sentiment analysis; $0.50 for entity sentiment analysis
> 20M units	Contact sales representative

6. Comparison

The following table summarizes the sentiment analysis offering from the public cloud providers. The pricing information has also been normalized to base on a unit of 1,000 characters for comparison.

	Amazon Comprehend	Microsoft Text Analytics	Google Natural Language API
Overall Sentiment analysis	✅	✅	✅
Entity sentiment analysis	❌	✅	✅
Accuracy	Datasets dependent	Datasets dependent	Datasets dependent
Text input limit (bytes)	5,000	5,120	1,000,000
Number of text languages support	12	15 for overall sentiment 1 for entity sentiment	16 for overall sentiment 3 for entity sentiment
Text language: English, Spanish, French, German, Italian, Portuguese, Japanese, Korean, Chinese	✅	✅	✅

Text language: Arabic	✅	❌	✅
Text language: Hindi	✅	✅	❌
Text language: Dutch, Turkish	❌	✅	✅
Text language: Norwegian	❌	✅	❌
Text language: Indonesian, Thai, Vietnamese	❌	❌	✅
Programming: REST API, CLI, Java, Python, .NET, JavaScript, Go, Ruby	✅	✅	✅
Programming: PHP	✅	❌	✅
Pricing/1K units: Up to 5K units (1,000 characters per unit)	$1.00	Free	Free
Pricing / 1000 units: 5K - 0.5M units	$1.00	$1.00 for overall / entity sentiment analysis	$1.00 for overall sentiment analysis $2.00 for entity sentiment analysis
Pricing / 1000 units: 0.5M - 1M units	$1.00	$0.75 for overall / entity sentiment analysis	$0.50 for overall sentiment analysis $1.00 for entity sentiment analysis
Pricing / 1000 units: 1M - 2.5M units	$1.00	$0.75 for overall / entity sentiment analysis	$0.50 for overall sentiment analysis $1.00 for entity sentiment analysis
Pricing / 1000 units: 2.5M - 5M units	$1.00	$0.30 for overall / entity sentiment analysis	$0.25 for overall sentiment analysis $0.50 for entity sentiment analysis
Pricing / 1000 units: 5M - 10M units	$1.00	$0.30 for overall / entity sentiment analysis	$0.25 for overall sentiment analysis $0.50 for entity sentiment analysis
Pricing / 1000 units: 10M - 20M units	$0.50	$0.25 for overall / entity sentiment analysis	$0.25 for overall sentiment analysis $0.50 for entity sentiment analysis
Pricing / 1000 units: 20M - 50M units	$0.50	$0.25 for overall / entity sentiment analysis	Contact sales representative
Pricing / 1000 units: > 50M units	$0.25	$0.25 for overall / entity sentiment analysis	Contact sales representative

Comparing the technology, while all these cloud services perform *polarity analysis and emotion analysis* to analyze the sentiment, the sentiment results are only represented in terms of polarity analysis (i.e. positive, negative, neutral, etc). That said, *Microsoft Text Analytics* and *Google Natural Language API* are more advanced as they support *overall sentiment analysis* as well as *entity-based (or aspect-based) sentiment analysis*, while *Amazon Comprehend* only supports *overall sentiment analysis*. Since the accuracy of these services vary on the datasets, companies who are interested in these services should further evaluate the sentiment detection accuracy of these services based on specific needs.

Comparing the text input support, *Google Natural Language API* clearly wins out by supporting long text up to 1,000,000 bytes, while *Amazon Comprehend* and *Microsoft Text Analytics* both

support short text input close to 5,000 bytes. The text length could be a critical factor for some companies to consider in their use cases.

Comparing the text language support, all these cloud services support many popular languages in the U.S., Europe, and Asia. While *Microsoft Text Analytics* and *Google Natural Language API* supports several more languages in total compared to *Amazon Comprehend*, *Amazon Comprehend* is the only one currently supporting both Arabic and Hindi which each is spoken by over 4% world population. The text language support is another important factor for some companies to consider in their use cases.

Comparing the programming languages support, all these cloud services support popular programming languages such as Java, Python, .NET, JavaScript, Go, and Ruby. Both *Amazon Comprehend* and *Google Natural Language API* supports PHP in addition.

Comparing the pricing, *Microsoft Text Analytics* is attractive for many use cases as it charges the same low price for both *overall sentiment analysis* and *entity sentiment analysis*. The pricing tiers of *Microsoft Text Analytics* are highly competitive and most of its tiers are often the cheapest compared to its competitors.

7. Conclusion

Cloud services for sentiment analysis are some of the most exciting and emerging offerings from the public cloud providers in the last few years. Each of these cloud services has its own strengths. As outlined in this survey, the level of sentiment analysis required (i.e. overall sentiment vs entity sentiment), the sentiment detection accuracy with respect to the datasets, the text input requirements (i.e. text length and text languages), the programming languages requirements, and the pricing structure, are important factors to consider. Ultimately, which cloud service to use really depends on the specific use case and datasets involved.

8. References

1. [Amazon Comprehend: Determine Sentiment](#)
2. [Amazon Comprehend: Detecting Sentiment](#)
3. [Microsoft Azure: Text Analytics](#)
4. [Microsoft Azure: How to: Sentiment analysis and Opinion mining](#)
5. [Google Cloud: Natural AI Language](#)
6. [Google Cloud - Analyzing Sentiment](#)
7. [Google Cloud - Analyzing Entity Sentiment](#)
8. [Google Cloud - Natural Language API - Interpreting sentiment analysis values](#)