Introduction

This technology review document compares and contrasts Amazon Comprehend vs. Microsoft Azure Cognitive Services. It will focus on the Text / Knowledge mining services of these two leading cloud providers: the capabilities, differentiators and recommendations.

There is a treasure trove of hidden knowledge in unstructured data. Customer emails, support tickets, product reviews, social media comments represent insights into customer sentiment and preferences that can be put to work for your business. How do you get at it?

At the outset, let's briefly examine the capabilities of each of these technologies.

Amazon Comprehend

Amazon Comprehend is a natural language processing (NLP) service that uses machine learning (ML) to help you uncover the insights and relationships in your unstructured data. For example, you can identify the feature most often mentioned when customers are happy or unhappy about your product. You can also use AutoML capabilities to build a custom set of entities or text classification models that are tailored uniquely to your organization's needs.

Amazon Comprehend Medical is used for extracting complex medical information from unstructured text, such as medical conditions, medications, dosages, strengths, and frequencies from a variety of sources like doctor's notes, clinical trial reports, and patient health records and the relationships within.

Amazon Comprehend provides APIs so you can easily integrate NLP into your applications. You simply call these APIs in your application and provide the location of the source document or text (such as S3 bucket). The APIs will output in a JSON format, which you can use in your application. Some of the key APIs are described below:

- **Sentiment Analysis**: overall sentiment of a text (Positive, Negative, Neutral, or Mixed).
- **Syntax Analysis**: identify word boundaries and labels like nouns and adjectives within the text.
- **Entity Recognition**: named entities ("People," "Places," "Locations," etc.)
- **Topic Modeling**: identifies topics from a collection, and maps to each document.
- **Custom Entities**: Using AutoML, identify terms that are specific to your domain.
- **Custom Classification**: build custom models using your business-specific labels without learning ML. For example, categorize requests by problem type based on customer description of issue.
- **Language Detection**: identifies text written in over 100 languages and the dominant language.
- **Multi-language**: text analysis on English, French, German, Italian, Portuguese, and Spanish texts.

Amazon Comprehend is fully managed: no servers to provision, no machine learning models to build, train, or deploy. You pay only for what you use: no minimum fees, no upfront commitments.

Azure Cognitive Services

Azure Cognitive is a broader set of services focused on more than just text. It spans the following:

- Decision: Make smarter decisions faster.
- Language: Extract meaning from unstructured text. Includes Immersive Reader, QnA Maker, Language Understanding, Translator, *Text Analytics*
- Speech: Integrate speech processing into apps and services.

- Vision: Identify and analyze content within images, videos, and digital ink.

Out of the above, the Language service (specifically Text Analytics) maps to Amazon Comprehend, hence e will focus only on Text Analytics.

Text Analytics helps discover insights in unstructured text using natural language processing (NLP), with no machine learning expertise required. It identifies key phrases and entities to understand common topics and trends. It classifies medical terminology using domain-specific, pre-trained models, extracts insights from unstructured clinical documents such as doctors' notes, electronic health records, and patient intake forms using the health feature of Text Analytics and determines relationships.

- Broad **Entity Extraction**: important concepts in text and named entities such as people, places, and organizations. Detect and extract 100+ types of personally identifiable information (PII), including protected health information (PHI), in documents.
- Powerful **Sentiment Analysis**: Discern what customers are saying about your brand, and detect sentiment around specific topics or features.
- Robust **Language Detection**: Evaluate text input in a wide range of languages and dialects.
- Key **Topics**: phrases that best describe the subject of each record using key phrase extraction.
- Privacy & Security: Your data stays yours. Microsoft doesn't use your training data.

You can run Text Analytics wherever your data resides: cloud, on-premises, or the edge in containers. Backed by Azure infrastructure, Text Analytics offers enterprise-grade security, availability, compliance, and manageability.

Comparisons

As we can see, Azure Cognitive is much broader than Amazon Comprehend. There are many similarities between Amazon Comprehend and Text Analytics of Azure Cognitive Language service: sentiment analysis, entity recognition, language detection, topic modeling and extracting medical information.

Amazon Comprehend seems to be more customizable with its custom entities and classifications. Azure Cognitive Text Analytics seems more flexible in terms of deployment options.

Additionally, I came across this sentiment analysis comparison based on sample data set from Kaggles.com and Google datasets measuring accuracy, speed and price. The accuracy of Amazon Comprehend vs. Text Analytics varied based on the data sets. However, Amazon Comprehend had a slightly faster response time and lower cost.

Conclusion

Amazon Comprehend and Azure Cognitive are state-of-the-art technologies providing competitive and compelling capabilities, which continue to increase. At this point, Comprehend seems to have a slight edge regarding customization and cost. However, your choice will also largely be determined by your current environment and investments made. In the end, you can't go wrong either way!

References

https://aws.amazon.com/comprehend/

https://azure.microsoft.com/en-us/services/cognitive-services/