Parabole

Beyond text analytics: The Parabole Way

SMRATI HANS

Abstract

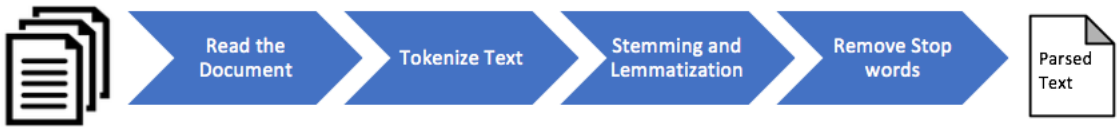
There are myriad software companies in the marketplace providing solutions that in whole or in part are based on text analysis. This paper discusses techniques used to process voluminous information in the form of text, images, PDFs, or any document utilizing Text Analysis and Semantic or Sentiment Analysis ("SSA"). We intend to demonstrate how SSA is rapidly eclipsing Text Analysis and how this technology can reduce not solely the time required to process information but increase, greatly, the quality and usefulness of the output. This advance has significant, even game-changing, ramifications across the enterprise.

Problem Definition

The Internet is relentlessly flooded with information and any analysis of value to the enterprise requires abilities to filter this massive and ever-growing content store. This information in this content store must be filtered, absorbed, and then converted into structured or semi-structured knowledge so that it can be processed and used within a specific domain. However, given the volume and complexity of this information, manual processing of these volumes is inefficient and not an alternative. Currently, some Text Analysis can aid in the processing of text and filtering data from within information. The next significant challenge stems from compounding of the N-dimensional analysis of this data with appropriation of its contextual references, that is, how specific terms present in the text can be related to other concepts and how one concept can impact other concepts. This is only possible if the analyzer, be it system or human, has adequate knowledge of the domain. It is seemingly impossible for a human to analyze the meaning of high-density information and derive all possible references of impact. For this paper, we will focus on the solutions developed at Parabole to solve this challenge.

Text Analysis

Text or Syntax Analysis is a pre-requisite for Semantic Analysis in extracting data, reshaping it, exploring it and visualizing it. The ultimate output generated from this primary process acts as input to machine learning algorithms whose output is true Semantic Analysis. By way of background, Text Analysis is generally performed using various and widely available Natural Language Processing techniques. The process of Text Analysis can be depicted as:



Semantic & Sentiment Analysis

With the now "parse" text, we may now proceed to Semantic or Linguistic Analysis which is directed towards **understanding the meaning** of the text in each document. The parsed text is represented in a multi-dimensional vector reference frame. As the volume of data being processed is ever-increasing, more knowledge is required in order for machines to achieve a level of understanding equivalent to a Subject Matter Expert. Due to this expansion, there exists the need for Feature Extraction & Transformation (FET) in order to reduce the dimensionality of information and increase the usefulness of

the underlying meaning of the information. At Parabole, this semantic analysis is accompanied by sentiment analysis or "opinion mining," as our focus is document context and individual paragraphs within documents, easing document analysis for financial institution personnel.

Semantic analysis is divided into two phases: the learning phase and the analysis phase. The overall process embeds techniques comprised of unsupervised learning, which includes Clustering (example: hierarchical clustering or k-means clustering), dimensionality reduction, as well as supervised learning (example: regression techniques and decision trees). These techniques are further supplemented by algorithms including neural networks, Latent Dirichlet Allocation (LDA), and Latent Semantic Analysis (LSA), etc., for overall calibration of the output.

Sentiment analysis is performed using a hybrid approach, leveraging machine learning techniques along with a knowledge graph, which is refined by deploying the knowledge of domain experts. This knowledge is represented in form of a triple-database (TDB) or an ontology which captures relevant information in the form of concepts, their weights, along with the hierarchies and linkages among them.

Summary

In a never-ending process, regulators issue new rules and update existing regulations; financial institutions are forced to comply within, and over, a specific timeframe. At Parabole, we analyze the risk and regulatory information from documents issued by regulators and agencies and subsequently, complete a deep-dive comparative analysis with internal documents, including procedure, model and compliance documents, in order to analyze the areas of impact of existing and newly issued regulations and rules. Our analysis is performed using our complete process of Text Analysis along with Semantic/Sentiment Analysis.

The benefits of using Text Analysis are:

- Reduced time required to read documents and convert unstructured data to structured information.
- Increased scope without requiring additional manual (human) resources.

Text Analysis alone assists professionals by producing structured information, however, to gain understanding and to process this information still requires expert analysis and full understanding of the domain. By enriching the text layer with Semantic & Sentiment Analysis, Parabole creates a knowledge graph from the information ingested, producing the following benefits:

- Reduced labor hours required to read and analyze documents which earlier consumed 80-90% of the total project time, this is now achievable in ≤ 20% of status quo in most organizations.
- This process can increase the dimensions in which processing can be achieved by utilizing machine knowledge
- Meaningful reduction in error rates relative to current manual methods.

Hence, Text Analysis is, simply put, not robust enough to develop a cognitive platform, it must be accompanied with Semantic and Sentiment Analysis in order to solve domain centric use cases.

About us

Parabole is a Princeton, NJ based cognitive analytics company, automating the creation of enterprise knowledge from unstructured sources of information. We provide our clients with platform to solve risk, finance and regulatory compliance-related challenges that depend on the knowledge-data interchange. We accomplish this by delivering a range of bespoke applications in the areas of credit, market, liquidity risk and data governance domains.

To learn more, visit www.parabole.ai OR reach out to info@parabole.ai