



INTERNATIONAL
INSTITUTE FOR
ANALYTICS™

DISCUSSION SUMMARY
RESEARCH & ADVISORY NETWORK



CLIENT ONLY

Text Analytics: Unlocking the Value of Unstructured Data

SIMRAN BAGGA

Senior Product Manager for Text Analytics, SAS

JULY 2016

Interviewed by Robert Morison, IIA Lead Faculty



DISCUSSION OVERVIEW

Most organizations - businesses, government agencies, non-profits - are just scratching the surface of what they can learn and accomplish through the analysis of unstructured text. Opportunities abound, applications are multiplying, and the landscape is coming into sharper focus. IIA spoke with Simran Bagga, Senior Product Manager for Text Analytics at SAS to get her views on how to unlock the value of unstructured data.

What can Text Analytics do for an organization - what is a good starting point?

Before outlining options for where and how an organization can start unveiling opportunities with unstructured text, it is important to understand what Text Analytics is and why one should care. At a high level, Text Analytics is about deriving information from text sources – client interactions, product reviews, call center logs, emails, blogs, tweets, and other forms of electronic text – so organizations can make business decisions more effectively. Text Analytics can reveal insights about the firm’s products, the products of competitors, and other information to improve business operations and performance.

In practical terms, the objective is to provide structure to unstructured data, essentially turn text into data for further analysis. Common applications include automatically categorizing text to organize large numbers of documents and mine that data, incorporating text along with other structured data for predictive analytics, listening to the “voice of the customer” – or citizen – and the sentiment behind what’s being said, and making business information searches and queries more intelligent by folding in relevant context.

Gartner’s IT Glossary defines Text Analytics as the process of deriving information from text sources for purposes that include summarization, classification, investigation, sentiment analysis (the nature of

commentary on a topic), and explication (what drives that commentary).

It’s essential to understand how this definition translates into action and the value you can generate through the various capabilities of Text Analytics. These include search and information retrieval, information extraction through techniques such as natural language processing (NLP), tagging or annotation, lexical analysis to study word frequency and distribution, singular value decomposition (SVD), pattern recognition, data mining techniques including link and association analysis, predictive analysis, segmentation, and visualization.

Simply put, without the core capability of converting unstructured text into a structured form, text cannot be analyzed in sophisticated ways. The alternative is a painstakingly manual and error-prone process.

How can Text Analytics be applied to solve today’s business problems?

Some of the most powerful applications are in customer service and experience. By analyzing contact center and other voice- or text-based interactions, organizations can understand what customers like and don’t like. They can determine the drivers behind customer behavior and anticipate customer needs. They can get to the root causes behind customer complaints and have an early-warning system for



product and service problems. With streaming technology enabling on-the-fly analyses, organizations can serve customers, make real-time recommendations to influence behavior, or even detect fraud at the point of interaction.

Analysis of social media content keeps an organization informed about what customers and others are saying about products, services, brands, and the company in general. All this customer intelligence can drive initiatives to reduce customer attrition, increase brand loyalty and net promoter scores, and design and improve the customer experience. Analysis of sales notes and other data can make revenue attribution more precise and reveal opportunities to up-sell and cross-sell.

We see more and more applications of Text Analytics across a variety of industries, for example:

- In health care, the management and interpretation of clinical notes is used to improve patient safety and care.
- In insurance and government, Text Analytics plays a growing role in fraud detection and investigation.
- In energy and manufacturing, Text Analytics is used to gather customer feedback for early detection of problems and to incorporate warranty issues into product design, resulting in cost savings, improved quality, and reduced repair rate.
- The legal profession is automating activities including contract analysis and classification to reduce manual effort and determine shared characteristics among contracts.
- The financial sector leverages text analysis to turn financial advisor notes into quantifiable measures of client experience, so they can better understand sentiment, identify clients at risk, and assess opportunities to deepen relationships.

By extension, Text Analytics can serve any function that wants to minimize the effort required to manage and organize large volumes of documents and/or wants to add value by mining and analyzing their content.

The analysis of unstructured text has been possible for years. What's different today - what can we do that we couldn't do before?

Demand for Text Analytics has skyrocketed. Forrester finds that Text Analytics implementations have doubled since 2012. Every organization, regardless of industry, has unmet needs and opportunities – and therefore growing interest in analysis of unstructured text. To complicate matters, there are new and rapidly emerging data sources all around us. These include the vast amounts of social media data, and lately the unstructured text generated by people's interactions with Chatbots and digital personal assistants like Siri, Amazon Echo, and Cortana. Data from those sources potentially represents the “voice of the customer”. Many of these applications that rely on social media entail knowing where the communities are, the language they use, and the trends and topics that interest them. The Internet of Things is also driving demand for applications that combine structured data such as operational details with unstructured data such as log files. As business people become aware of the possibilities and what that can mean for business performance, demand just continues to grow.

Demand is also driven by growth of supply, not only of data but of the technologies to manipulate it. There are lots of technology options for various types of analysis, including open source and cloud based tools, and they're getting easier to use. An enterprise can buy a full Text Analytics solution, or technology components to create their own platforms and applications, or very specific applications “as a



service.” An example would be an off-the-shelf social media based tool that can provide both sentiment analytics and competitive intelligence for the hotel industry. In addition to the application of machine learning to text mining, there are also domain-specific taxonomies available for a wide variety of business applications. And behind the scenes, faster processing capabilities and access to more and more data enable higher-quality NLP. Organizations can do more sophisticated predictive and descriptive analytics, which further whets their appetites.

Finally, what’s really different today is that we can do so much analysis in real-time and in-stream. Analysis of large data sets used to be a batch process with associated delays. With today’s big data analytics technologies, we can process data as it comes in. We can analyze customer sentiment and preference, categorize or score the customer to predict behavior, and recommend what to do next – all in real-time while engaged with the customer. The speed of analytics has been a game-changer.

Who are the main players involved in developing and using Text Analytics solutions?

For advanced solutions, a variety of roles and capabilities may come into play. Let’s start with the **decision maker**, the person who applies the results of Text Analytics to work more productively and to make faster and more informed decisions. This might be a business or product owner, chief customer officer, director of risk management, or head of investigations. These people want the analytics to fit into their workflows and work with the output, drill-down and explore and understand patterns.

Another player is the **domain expert** or business analyst, who combines in-depth understanding of the application domain with the insights from Text Analytics to integrate pertinent information into a

coherent message. For example, a law enforcement analyst or investigator wants to solve crimes more quickly by connecting the dots from one crime to another, to find the linkages and patterns within the narratives of criminal records without having to read through hundreds or thousands of records.

A key role, specific to Text Analytics, may be a **linguist** who is expert in writing linguistic rules and defining corporate taxonomies and lexicons. In practice, many organizations have to compromise on this expertise and, depending on the use case, others fill this role in some capacity. Many have seen success with the linguistic responsibility being picked by a business expert or analyst if a domain-specific taxonomy is available and the goal is primarily text extraction, exploration, categorization, or sentiment analysis, and not using unstructured text in predictive models.

Data scientists build the machine learning and statistical models that can be deployed in batch or real-time operational systems. They work with domain experts and business owners to create “gold standards” of documents or other content which can be used as the basis for Text Analytics models. And they provide checks on the statistical validity of the decisions being made. With the right technology and training, the data scientist can develop skills in computational linguistics, which is important to understand and interpret text parsing and related functions.

As you think about this cast of characters, keep in mind the interactive and iterative nature of Text Analytics. To enable these individuals to collaborate regularly and effectively, it’s important to equip them with tools and technology that provide value out of the box. Our goal is always to take the complexity out of the text analysis process, while still creating a flexible and scalable environment for a variety of solutions.



What are some of the challenges organizations face when implementing Text Analytics solutions?

For many organizations, the biggest challenge may be picking their shots and focusing their energies. There are so many new opportunities and options for employing Text Analytics that people are asking where to start or what to do next. They need to understand what's possible given their combination of business need, available data, and the people, processes and technology to put text data to work. They should also explore how to blend analysis of unstructured text with other analytics and existing applications to create more robust business solutions.

There's a parallel challenge on the technology side with the growing number of tools and techniques available. What technological capabilities does the organization need? You can put together some basic Text Analytics capabilities, but then can your platform scale and interface with all your data sources and applications? There is no one-size-fits-all Text Analytics technology on the market, so organizations do well to implement software that is flexible enough to address a variety of business use cases and support a variety of user personas, each using Text Analytics to make data-driven decisions in different ways.

Also be alert to skills gaps. Data scientists are in short supply to begin with, and there may be an unrecognized gap in the linguistic science talent. Many universities now offer statistics and analytics programs, but very few teach and practice text analysis. Many young data scientists – statisticians, mathematicians, economists – do not readily see Text Analytics as an area of opportunity. This can pose challenges for organizations when getting started with text projects – they need strategies for building, maintaining, and supplementing their capabilities. It helps to have technology and services partners who can fill gaps as needed.

Global organizations have special considerations and challenges around whether to do Text Analytics in a variety of native languages and then combine the results as needed, or to translate all text into a common language for analysis – and perhaps lose something in translation.

Finally, everyone faces the everyday challenges of screening and preparing text data for analysis – stemming terms, fixing misspellings, resolving acronyms, extracting custom entities and facts, understanding slang, perhaps even recognizing sarcasm.

What makes Text Analytics unique and exciting, especially as you look to the future?

The excitement starts with business results. More and more organizations are seeing the value and opportunity in Text Analytics, seeing how the analytics accelerate interpretation and insight in a wide range of business domains. People don't have to read all the documents and notes. They receive the essence of what the text holds, and then they can dig in at will. It's never easy to alter a business process, but it may be easier with Text Analytics because of the enormous boost in productivity and cycle time to insight.

Other things make this work exciting. The textual data is constantly changing, especially when it represents conversations with customers. With Text Analytics, we can spot deeper trends quicker. The business owners and analysts are constantly learning and refining not just how to interpret the text, but how to improve the conversation. Text gets businesses closer to the events and opinions that are relevant today. The data often has been created directly by customers or by relevant stakeholders in a process – and that adds a dimension to understanding reality that we have not had in the past.



We'll continue to see technological advances. Optical character recognition (OCR) technology has turned our paper documents into electronic ones. Today's frontier for many organizations is voice-to-text transcription. So the available relevant data keeps growing, yet we can still process it more quickly than ever, including in-stream. Perhaps the most exciting advances are in the outputs of analytics, as systems get better at continuous learning, offering options or assistance, explaining their reasoning, and interacting more like a human.

The ability to incorporate human input and domain specific rules is what makes Text Analytics unique, even in comparison to other forms of analytics. The most sophisticated forms of cognitive computing may be at work, but they require Text Analytics for natural language interpretation and analysis. The key here is to take away the complexity from the users, while providing the much needed transparency into the analysis process so users can understand the outputs and recommendations.

What are the top three things any organization should know or do when adopting Text Analytics?

First, start with an end in mind. What are your goals, both broadly in the business domain and for the specific analytics initiative? Start with the business objective and then look for opportunities to analyze data previously ignored because it was unstructured. A well-defined objective will tie into the following questions: How do you want to deploy the analytics? Who are the consumers of the results, how will they interpret the analysis, and what decisions will they make? For example, what are the top priorities for the customer experience that you want to deliver? Are they customer service, then product quality, and then price? Some Text Analytics methods, including sentiment analysis, are often called into question because they're applied to the wrong problem.

Second, define your success criteria. What does it mean to reach the goals? Success may be expressed in terms of improving a specific business performance metric through an analytics model of specific accuracy. A clear definition of success helps the analytics team find the most direct path. For example, if success means increasing customer satisfaction by a certain percent by accurately categorizing call center notes or service records into relevant groups, the analysis can perhaps be highly automated with minimal impact to operational systems and processes. However, if success means injecting analytics into call center operations to make real-time recommendations using text and predictive analytics, this will be a more resource-intensive effort involving process change that should be well worth the effort.

Third, start small and think big. Your organization may get its feet wet with modest or off-the-shelf applications of Text Analytics, but plan for the future; have a business vision and a capability expansion strategy. There are questions that have gone unanswered because they were buried in unstructured text, and now technology is not the limitation. Do your due diligence and research on Text Analytics capabilities and the pros and cons of different approaches. Anticipate how data sources and volume and movement will grow. Also anticipate how you will scale Text Analytics technologically and perhaps globally. Be realistic about filling your talent and technology gaps, and get help where needed from trusted advisors.

The opportunities with Text Analytics will continue to multiply. The business demand for Text Analytics will grow fast. How will you prepare to meet the demand and realize the opportunities?

Additional Information

To learn more about this topic, please visit www.sas.com/textanalytics



About the Interviewee



SIMRAN BAGGA

Simran Bagga is the Product Manager for Text Analytics at SAS Institute. Her background is in advanced analytics, and she has many years of hands-on experience with predictive analytics, data mining, Text Analytics, model management and analytics deployment into operational systems. For the past four years at SAS, she has helped customers in many industries to understand the value of SAS Analytics including Text Analytics and how it can be plugged into their enterprise to solve specific business problems. Earlier she worked at IBM SPSS in a similar capacity. Prior to joining SAS in 2012, she managed an analytics team at the Mayo Clinic with the goal of standardizing their analytics and BI practice.

IIANALYTICS.COM

Copyright © 2016 International Institute for Analytics. Proprietary to subscribers. IIA research is intended for IIA members only and should not be distributed without permission from IIA. All inquiries should be directed to membership@iianalytics.com.