

April 2016

Decisions, Decisions

Loretta Chan | Mary DeBisschop

A man in a suit is looking through a magnifying glass at a complex data visualization. The visualization includes a bar chart, a line graph, and a table of data. The background is a solid blue color.

Which text analytics
tool is right for you?

Decisions, Decisions

Loretta Chan | Mary DeBisschop

The opinion is now widely shared that the ability to extract actionable insights and intelligence from unstructured text is crucial for most organizations. Your data might include verbatim feedback from customers, prospects, employees, or members of the public, for example.

So, you have your unstructured data in-hand. Now what? How do you get started? There are so many tools out there – which do you use?

Venturing into the world of text mining and analyzing unstructured text can sometimes feel like walking into a car dealership – you are faced with an array of vehicle models, each with its own set of options. When confronted with these types of choices, you are likely to make trade-offs as you attempt to get the most value for your money.

Each of the available text analysis solutions also presents different strengths, challenges, and functionalities. Depending on factors such as the characteristics of the data you are looking to analyze, requirements such as set-up cost and turn-around time, and how you want to share the results, some technologies and tools might be better suited than others.

Although some hybrid options exist, the main text analysis tools can be placed into one of the following three categories:

- **Manual Coding**
- **Linguistic/Rule-based Text Analytics**
- **Machine Learning Text Analytics (Supervised & Unsupervised)**

Manual coding is straightforward in that 100% human involvement is required throughout the entire process from defining, identifying, extracting, and classifying concepts within unstructured text. Historically, analysts have been performing this rather labor-intensive task of manual coding, which typically entails multiple iterations of reading through comments, mentally flagging recurring themes, grouping similar themes together to form categories and sub-categories, and then classifying each comment into the most relevant categories and/or sub-categories.

DECISIONS, DECISIONS

April 2016



Loretta Chan | Mary DeBisschop

Linguistic/Rule-based text analytics relies on both the human analyst and a linguistics-driven engine to work together. The engine automatically segments a comment into fragments (often based on parts-of-speech), which are then tagged, grouped based on similarity, and indexed before being extracted as concepts. The human analyst defines categories and creates rules to classify extracted terms and concepts.

Machine Learning Text Analytics can be supervised or unsupervised. Machine learning options incorporate and rely heavily on statistical or mathematical algorithms to aid in the classification process. Supervised machine learning is similar to linguistic/rule-based methods in that human interaction is essential in the text classification phase. Unsupervised machine learning is just that... it relies heavily on the algorithms to classify data, with very minimal or no human involvement.

Ipsos' extensive experience in working with clients on unstructured data analysis has led us to the conclusion that the expertise and guidance of an analyst is needed to choose the right tool for your particular situation. The analyst will ask you several questions about your requirements and will also examine the dataset in order to determine the ultimate path of the analysis.

On the right and on the following page are several of the key factors to consider when making a choice of text analysis tools. The table that follows displays the appropriateness of each tool given each factor.

1. Set-up labor, analysis turn-around time

Since 100% human involvement is required with manual coding, it's understandable that it tends to require the highest amount of set-up labor and that the turn-around time for analysis is the longest vs. other text analytics tools that utilise algorithms and/or technology.

2. Big data friendly, scalability, consistency/replicability

These factors speak to the ease of processing large amounts of data, data from multiple data sources, and the ability to replicate the analysis while maintaining consistency even as the data set grows. In fact, scalability is at the heart of the benefits of text analytics tools that utilise some type of technology. A high volume of comments ensures that non-manual text analytics approaches remain time and cost effective over time. The definition of "high" volume will depend on the tool you choose, but as a rule of thumb, several thousand or more comments are required to realize the true benefits of non-manual text analysis tools.

3. Ongoing human involvement

Ongoing human involvement is typically needed to fine-tune the coding or text modeling. In addition to the obvious cost implications, the amount of ongoing involvement required in utilizing each text analysis tool should be considered in planning for other project management-related needs, such as resource allocation, as well as any ongoing training of new resources.

DECISIONS, DECISIONS



April 2016

Loretta Chan | Mary DeBisschop

4. Transparency

Machine learning algorithms often feel like a “black box” as they are less transparent in terms of the inner-workings of the text analysis. This can be particularly troublesome when output or results from the analysis are counter-intuitive or nonsensical, as it makes them more difficult if not sometimes impossible to reverse-engineer, adjust and explain.

5. Flexibility, level of control

Following from the previous point, when results are counter-intuitive, the analyst will want to intervene and make improvements to the text analysis algorithm or model. In particular, unsupervised machine learning options tend to offer the least amount of flexibility and control since these tools depend on the algorithms used, without any up-front human intervention or training (hence the term “unsupervised”).

6. Capture hidden patterns/identify latent dimensions

Creating rules within a text analysis tool often involves the analyst inputting explicit references and specifications. It is very challenging for linguistic and rule-based text analytics tools to deal with abstract or latent (less prominent) concepts. Both supervised and unsupervised machine learning algorithms are better than linguistics and rule-based options on this front, as machine learning techniques are designed to look at both the explicit and latent dimensions in the data.



DECISIONS, DECISIONS

April 2016



Loretta Chan | Mary DeBisschop

	Manual Coding	Linguistic/ Rule-based	Supervised Machine Learning	Unsupervised Machine Learning
Set-up labor	★	★★	★★★	★★★★
Analysis turn-around time	★	★★	★★★	★★★★
Big data friendly/Scalability	★	★★★★	★★★★	★★★★
Consistency/Replicability	★★	★★★	★★	★★
Ongoing human involvement	★	★★	★★★	★★★★
Transparency	★★★★	★★★	★★	★
Flexibility/Level of control	★★★★	★★★	★★	★
Capture hidden patterns/ identify latent dimensions	★★	★	★★★	★★★

★★★★★ Better/Positive —————> ★ Worse/Negative

CONCLUSION

After going through the list of key factors to consider when making a choice of text analysis tools, it is easy to conclude that the analyst is key in setting the text analysis on the right path, carrying out the text analysis in the right way, and interpreting, validating, and contextualizing the results to the individual business questions. The analyst is at the heart of the process in most of the projects that Ipsos runs for clients; technology is an enabler but will not give you all the answers. We recommend that you discuss each of the factors above with your analyst in order to set your text analysis effort up for success.

DECISIONS, DECISIONS

April 2016

Loretta Chan is Vice President for US Loyalty, Advanced Analytics, Ipsos.

Mary DeBisschop is Senior Vice President for US Loyalty, Financial Services, Ipsos.

And a special thank you to **Jean-François Damais** and **Fiona Moss**

Ipsos Loyalty is the global leader in customer experience, satisfaction and loyalty research with over 1,000 dedicated professionals located in over 40 countries around the world. We collect, integrate, analyze and model a vast amount of both unstructured and structured data on a regular basis. Our modeling & analytical expertise, technological capabilities, and EFM partnerships enable us to offer creative solutions that build strong relationships and lead to better results for our clients. This has made us the trusted advisor to the world's leading businesses on all matters relating to measuring, modeling, and managing customer and employee relationships.

This *Ipsos Views* white paper is produced by the **Ipsos Knowledge Centre.**

www.ipsos.com
@_Ipsos

GAME CHANGERS

<< Game Changers >> is the **Ipsos** signature. At **Ipsos** we are passionately curious about people, markets, brands and society. We make our changing world easier and faster to navigate and inspire clients to make smarter decisions. We deliver with security, simplicity, speed and substance. We are Game Changers.

GAME CHANGERS

