GUIDEBOOK
SPSS TEXT MINING

December 2007
THE BOTTOM LINE

Text mining can help companies leverage all the unstructured information they have about products, services, competitors, and customers to increase customer satisfaction and loyalty. SPSS Text Mining leverages SPSS’s data mining platform to enable companies to rapidly analyze unstructured information for better decision making.

THE SITUATION

Organizations are increasingly challenged to measure and maintain customer loyalty. While many companies develop internal programs to develop and measure loyalty, the bigger challenge is the multiple sources of information customers have about other customers’ experiences, competitive services and offerings, and others’ commentaries on the value they provide. Although some companies conduct customer surveys today, the ability of surveys to deliver real insight are limited by our ability to move beyond yes-no and multiple choice questions to analyze the real satisfaction of customers — which often is relayed by open-ended survey questions, comments recorded in a call center call, and beyond the enterprise in user blogs and discussion groups on Web sites, new stories, and other text-based communications.

The next challenge in retaining customers, driving new sales opportunities, and understanding how to best manage customer relationships is in moving beyond transactional and basic survey data to understand all the communications about products, services, and customer satisfaction in context.

One customer said, "The key benefit from text mining is the insight into the customer opinions you can’t get from other data: opinions about our service; problems we need to fix."

However, putting all the streams of unstructured information in context is simply not cost-effective without text mining. Text mining, either on its own or in conjunction with traditional transactional data mining, can help companies put customer insights in context, identify areas for improvement, and build more sophisticated models to understand, predict, and proactively manage customer interactions — and customer loyalty.

SPSS Predictive Text Mining enables users to analyze, categorize, and draw conclusions from unstructured data such as text.

Key components of the solution include the SPSS text mining extraction engine and the Clementine workbench.

Text Mining enables users to extract knowledge from unstructured text data, by identifying core concepts and trends that can be analyzed for better business decision making. The SPSS Text Mining engine enables you to extract key concepts, sentiments, and relationships from textual or "unstructured" data and convert them to a structured format that can be used to create predictive models.
It can process common document types as plain text HTML, XML, PDF, and Microsoft office documents. The extraction engine can also read, mine and synthesize content from databases, RSS feeds and blogs. It natively supports English, French, Spanish, Italian, German, Dutch, Portuguese, and Japanese languages; many other languages are supported through translation. The text mining engine can extract domain-specific concepts such as abbreviations and acronyms, identify and extract items such as addresses or social security numbers, and identify and categorize sentiments — such as likes and dislikes — based on its opinion dictionary.

The Clementine workbench enables users to build predictive models using both structured data and text in a visual workflow environment, so users can apply classification, clustering, and other modeling techniques to both types of data.

This guidebook, based on Nucleus Research’s analysis of SPSS text mining customers, identifies the key best practices, missteps to avoid, and fine-tuning tips that companies can utilize to integrate text mining into their analysis to better understand customers.

Customers analyzed included companies in the financial services, telecommunications, high technology, market research, public and nonprofit, and automotive sectors in the United States and Europe.

In its analysis of companies using SPSS Text Mining, Nucleus found key benefits achieved included:

- Reduced churn. Organizations such as financial institutions and communications service providers can leverage call center and other information to identify better programs to retain profitable customers.

- Improved management of promotions. Better predictive modeling that leverages unstructured as well as structured information can increase marketing and promotional campaign results.

- Increased visibility. The ability to actually digest and understand key themes from focus groups, surveys, blogs, and other sources of commentary on a company's products and services can help them make better decisions.

- Increased analyst productivity. The ability to automate much of the initial coding of information, and the ability to use Clementine to structure and manage the analysis reduces the time needed to effectively analyze information.

Nucleus found some analysts were able to increase productivity by up to 50 percent.

- Improved product development and refinement. Rapid analysis of customer feedback can enable decision makers to refine and improve products to better meet customer needs.

Although many companies are at the early stages of leveraging text mining, they can provide some valuable insights on the future of driving better decision making through text mining. For example, one company is using text mining to analyze all
its customer feedback to identify when customers are dissatisfied with processes. It found it was able to dramatically increase customer satisfaction by using insights from comments to change the customer experience.

Using text mining, a telecommunications company has seen 51 percent of its dissatisfied customers become company promoters (very satisfied customers) after two months.

Another telecommunications company has incorporated text mining into its overall data mining and promotions process, and has reduced customer churn by approximately 40 percent.

When properly implemented as part of an overall customer satisfaction strategy, Nucleus has found that text mining has the ability to cut customer churn in half.

BEST PRACTICES
Companies adopting text mining today have followed a number of best practices to effectively analyze text resources.

Move beyond data
Moving beyond analysis of transaction data to integrate text data can help analysts to build better models and provide more context around customer behaviors. The first step for many companies is to better analyze the existing internal text data, such as survey responses, customer letters, and call center records, that they have. Using Clementine as a common structure for modeling both traditional and text data enabled SPSS customers to build better models and use text information that wasn’t cost effective to analyze before:

- "Before, we’d have to choose: do we want qualitative research or quantitative? Now we are not having to make that decision because we can offer the combination."
- "We use Clementine as our primary modeling toolset. We started using text mining 2 years ago because we wanted to analyze text strings so we could tell if contacts were businesses, residents, estate agents, or other types of customers to put them in the right model to understand them better."

Pilot with your own content
Whether a company is considering SPSS or another text mining application, it shouldn’t rely on a demonstration of the applications’ capabilities with generic data. Nucleus found many companies had unique language and terminology that could impact the effectiveness of a text mining application.

Take advantage of training
Analysts shouldn’t assume that knowledge of Clementine and data analysis will make them instant experts with text mining. Understanding how dictionaries function, how the results of analysis are presented, and how text analysis can be integrated with other models is key to getting value from text mining software:

- One customer suggested, "Start with one of SPSS’s free Webinars – it helps you become familiar with how it works. Don’t try to figure it out on your own."
Other customers used the help of consultants in the beginning to learn how to get started.

- Another said, "It took us about three months to learn the software but I am not a heavy user. It’s pretty much automated so I haven’t done to any training other than poking around in the book myself."
- Yet another said, "Don’t get lost in trying to figure it out on your own. Go through a real tutorial where you can ask questions that are applicable to your own personal business objectives and address them with SPSS. That way you won’t waste time trying to figure out what you don’t understand that you may not need."

**Set realistic expectations**

Although text mining technology has made significant advances in the past few years, it’s important to note that it does not replace the job of the analyst — it simply does a lot of the initial heavy lifting around coding and categorization that can make analyzing unstructured information too expensive and time consuming to be beneficial.

**Use Clementine for data mining first**

Although many companies purchase data mining and text mining software at the same time, most take a phased approach to deployment. Companies start by learning the structure of Clementine for data mining and then integrate text into the model, using the Clementine structure to ensure repeatability and flexibility. Nucleus strongly recommends this approach, as it enables analysts to integrate text mining directly into the analytical process and improve the accuracy of predictions. The usability of Clementine can also shorten the learning curve to effective text analysis. As one user said, "One of the reasons we chose Clementine was because it can go against almost any data source. You can take the data and load it into tables, merge it with any format, and be able to do analysis. It has very good ETL capabilities so you can create a dataset from 10 or 15 data sources."

From a deployment perspective, users can deploy the entire data mining process automatically with the Clementine solution publisher runtime and send text mining extractions directly to Clementine to export to other SPSS predictive analytics applications.

**MISSTEPS TO AVOID**

As text mining is an emerging technology area, many companies find starting with a pilot project can help them to understand the power as well as the learning curve of using text mining. Companies using SPSS text mining identified a number of key missteps to avoid while adopting text mining.

**Don’t just focus on churn**

Although many companies have been effective using text mining to refine their customer churn models, text mining can also be helpful in planning customer acquisition and promotion models:

- One company, for example, uses text mining for time management, saying, "There’s so much information that we receive and don’t have time to read. Text mining lets us extract and summarize a 10-page document into one page
so we can read more press releases, transcripts, interviews, really anything than we possibly could without it.”

- Another company uses it to analyze customer service call information to determine what drives peaks in customer service calls and when. It has found many increases in calls are related to Web site related issues that it has now been able to correct.
- Another customer in the technology sector used text mining to quickly provide senior managers with feedback on a particular product line, saying, “We were dealing with a lot of qualitative feedback relating to systems integration for various products, and needed a quick and dirty tool because business leaders wanted to know what was going on.”

**Don’t ignore the complexity of the dictionary**

Although SPSS text mining technology includes standard definitions and terminology, most customers found there was some effort needed to augment or refine the text dictionary model to optimize results:

- “Training the dictionary is important. SPSS does deliver the initial setup but it’s not company specific, obviously. It can be different, for example, for a different call center with different pricing plans. You customize it for your business, which can take some time because even with the same company people use different synonyms. It’s an ongoing process but the first time it took us two or three weeks to really get things settled.”
- “The problem in healthcare is that the language is very complicated and often in the notes the spelling is very bad. We’ve focused on working with small sets of data with less complicated response categories.”

The technology natively supports English, French, Spanish, Italian, German, Dutch, Portuguese, and Japanese languages; many other languages are supported through translation. SPSS has developed a number of base dictionaries for industries including defense, telecommunication, and health care; however, it is likely some investment will be required to fine-tune the dictionary for a specific environment.

**Take time to prepare your content**

Just like with any data mining effort, some time will be required to prepare text information for effective mining and analysis. As one customer said, “The actual text and data mining is the smallest part. It probably takes 80 to 90 percent of time to prepare your data and then you do the mining.” SPSS Text Mining natively supports a number of different text sources, including Microsoft Office files, call center notes, survey text responses, e-mails, Web forms, and RSS feeds.

**Don’t expect text mining to do it all**

Although text mining has advanced significantly in the past few years, it still isn’t a complete substitute for manual analysis — however, it can be a valuable tool for helping analysts more effective. As one user said, “It didn’t comprehensively categorize everything, but it’s a time factor. We found the product was good at handling 40 percent of the analysis.”
FINE TUNING
As companies continue their journey with text mining, they find they can take a number of steps to leverage even more value out of the solution. Many of these fine-tuning tips are about leveraging new data sources and features, but equally important is gaining insights from peers and text mining experts.

Incorporate more sources
Once users become comfortable with the text mining application, they find they can identify and leverage new sources to build models that provide even more insight. On company, for example, started with a limited number of information sources and is now looking how to incorporate more information feeds into the model. Another research firm incorporated survey data, data from industry discussion boards on the Web, and industry newsletter to provide a fuller view of trends and themes in the industry.

Evaluate new features
As new versions of the software become available, companies using SPSS Text Mining should review both the features and user interface to determine how they may be able to achieve incremental ROI by upgrading. Some customers upgrading from previous versions found that improvements in the user interface made a big difference in the time needed to learn and use the application — particularly for casual users.

For example, one customer that is a Clementine user just beginning to implement text mining said, "We’ll continue to look at new releases moving forward. We have found that they come out with new releases and I’ve already seen that they’ve made a couple of changes to our products based on our suggestions — it seems like they’re responding to customer needs and try to get new features into their products. That’s a plus for us."

Introduce emotive techniques
Particularly in the call center area, leading-edge users of SPSS Text Mining are going beyond basic text categorization and analysis to use second-level analysis and emotive techniques to better understand customers and predict behavior:

- One customer said, “No one calls to say they’re terminating their contract. They ask what the early termination fee is, or about a different pricing plan. If someone calls all the time, we’re not too worried. If it’s a high value customer who calls once, that’s a real red flag for us.”
- Another said, "We’ve also been using emotive techniques so we can measure psychological attributes as well. Based on word choices, we have normative databases and we can see if certain emotional or psychological attributes are significantly higher. We can identify, for example, if someone is likely to have a need for affiliation, or a need for power."

Learn from your peers
Although text mining can often present a significant competitive advantage, Nucleus has found that many SPSS users are willing to share insights, techniques, and other tips as presented in this report because the evolution of text mining depends on more and more users gaining insights on how it can deliver value. Industry seminars, user events, Webinars, and training courses — as well as
developing relationships with peers that may or may not be in your vertical market — can help you identify new potential opportunities for text mining to deliver value.

**Continue to evaluate and evolve**
Finally, companies should keep in mind that as more sources of text data become readily available, as the volume of information grows, and as the sophistication of text mining improves, the best text mining strategy is one that continues to look for opportunities to leverage more insight from unstructured information. As one customer said, "A few years ago people said that they didn't have the data. Now you have data and say, I just need the time to analyze it. That's why we're continuing to expand what we're doing, because there are a lot of things we have yet to uncover."

**LOOKING AHEAD**
Today, two in three employees suffers from information overload — that is, they feel they have too much information they need to analyze to effectively perform their job (please see G75 – *Benchmarking Content Management*). That problem is only multiplied at the enterprise level as organizations try to integrate, analyze, and act on multiple and sometimes conflicting sources of insight on customers, products, and the competitive landscape. It is clear that text mining can deliver real benefits, both in terms of analyst productivity and in terms of broader business objectives such as reducing churn and improving product management and promotions.

For individual analysts, text mining can dramatically reduce the amount of time needed to deliver real insight, resulting in better decision making and improved product management.

Nucleus expects that as text mining becomes more standardized and easier to use, it will become a more common desktop tool for a broader user set.

At the enterprise level, integrating text analysis into a predictive analytics strategy can help companies first build richer and more complete models for analysis and secondly integrate those models into standard applications and processes, such as the call center or marketing campaign planner, to become more predictive and deliver even greater return on investment.