Leveraging Unstructured Text Data for Banks

Abstract

Banks have volumes of unstructured text data related to customers, businesses, processes and employee engagement which are a source of insights to generate more business, retain customers, improve productivity and enable effective decision making. This paper provides an imperative for adopting unstructured data mining by providing some use cases and examples. It delves into some unstructured data mining techniques that can be leveraged to address the bank’s business objectives.

The Imperative for Mining Unstructured Data

Analyst firm Gartner’s [1] report, CEO Survey 2012: Financial Services CEO Agenda, reveals that the global banking industry faces huge challenges in the form of dramatically reduced fee income and net interest income, with capital requirements increasing simultaneously. The report also predicts that in order to address these challenges, the prime focus of IT implementations should undergo a paradigm shift to ensure improved customer experience, build new revenue sources, enhance operational efficiency and predict and mitigate risk more efficiently. Unstructured text (data that does not necessarily fit into a single form, format or field) analytics can refocus IT implementations to help address these challenges -by offering insights to improve the three key functional areas of a bank i.e. performance management, risk management and customer relationship management. It can help identify opportunities for customer retention, business development and enhance efficiency of processes and employees.

Unstructured Text Analysis – A Strategy Driven Analysis

Banks have access to large volumes of unstructured text data from within organizational boundaries as well as the World Wide Web. This data relates to customers, employees, vendors, partners, processes and regulatory compliance. Advancements in Big Data technologies have now enabled banks to process vast amounts of unstructured text data to meet their various objectives- insights from such data can be used to understand customers, employees and competitors, and to improve the efficiency of existing processes. Figure 1 gives an overview of the ways unstructured text data can be used for business development.
The following sections describe how internal and external data can be used to derive insights for banking and financial organizations.

**Leveraging External Data:**

Customers today live increasingly digital lives, with access to technology and information that enables them to make better, more informed and more efficient decisions in their day-to-day interactions with companies and other individuals. Consequently, through a multitude of actions customers leave behind their digital footprints which can be used to understand them as never before. While banks have traditionally used peta-bytes of transactional data residing in their warehouses to understand customers, this data essentially gives them only a partial view. Transactional history does not necessarily capture the larger persona or the world-identity of customers which goes beyond age, sex and income. Factors such as hobbies, beliefs, emotional quotient, social standing and propensity to be influenced may ultimately dictate the customers' financial behaviour.

With an increasingly large number of people willing to adopt social media as a channel for conducting business, banks now have access to information about their customers over and above what is stored in their CRM systems. Gathering data from social media, however, does not automatically mean gaining effective insights. Social media being a large and unwieldy animal, it is difficult to distinguish between useful data, referred to as ‘signal’, and un-useful data, referred to as ‘noise’. Social media contains information in the form of unstructured text, images, interaction graphs and subscriptions to different groups, and so on. It is very important to understand the techniques that can be applied to a given type of data to extract meaningful information.

However, privacy and security concerns on social-media have prevented this from picking up pace in the banking sector. Thus, banks may have to use social-media as a source of broader or more generalized information about customer segments or communities rather than about individual customers. A better and more granular segmentation of customers will help banks in identifying the influencers and pushing targeted campaigns to customers. Banks may use unstructured text data for ‘skip tracing’ – a defaulting customer may have abandoned his physical address but is likely to continue with his virtual footprint in the social media, which can be used to trace the person – a known mechanism to trace credit card and other defaulters.

Other external sources of rich information for banks are media websites, online magazines, and blogs, available on the internet. These can be a source of competitor information – new products, services, campaigns and events that competition launches, announces or participates in. It is an open book that contains information about various strategies adopted by competitors as well as customer reactions to competitor products. This can also become a useful source for gauging wider perception about the bank's own products, services and brand. Open consumer forums often provide crucial information about issues that are not covered or captured in regular surveys and routine customer interactions. This information can subsequently be used to design better products, services and bring about operational efficiencies. The Web, in general, is a great source to understand the changing global population, its desires and expectations as well as gather knowledge about how to keep up with the rapidly evolving technological, financial and political landscape.

**Working with Internal Data:**

For banks, customer interactions on dedicated websites or discussion forums, and through emails, calls, complaint-logs or surveys offer another rich but unstructured collection of relevant data. Though unstructured, these sources are less noisy since they contain data generated from focused interactions and are often centered on a specific action. It is also easier to link information extracted from this type of unstructured content to specific transactions and therefore to specific customers. In an age when customers are steadily moving away from in-branch face-to-face interactions to net-banking or mobile-banking, effectively utilizing the information from all these sources and providing a seamless experience is the key to successful customer service management.

- Customer lifecycle evaluation
- Fraud assessment
- Systemic issue based customer pain point and service gap identification
- Inherent issue detection for products and services
- Cross and up sell opportunity identification
- Staff and branch performance feedback
Equally important is intra-organization employee data from interactions, mails, intranet, and internal social media such as blogs, forums and ESS (Employee Satisfaction Survey) to gather insights on:

- Capabilities and gaps, and hence up-scaling or training requirements
- Process compliance, for operational risk management and fraud prediction and
- Attrition or churn analysis

To gain useful insights from external and internal data, and to employ this to productive use, it is important for the banks to answer the following as seen in Figure 2.

Since majority of the data that banks come across is in the form of unstructured text, we will provide a view of the various steps involved in a text analytics life cycle which encompasses the gather, discover, deliver, pre-empt, to measure and evolve stages, as depicted in Figure 3.
Processing Text Data to the Bank’s advantage

It is evident by now that there is a big case for banks to adopt text mining techniques to gain useful insights for improving performance management, risk management and customer relationship management.

Sentiment Analysis and Opinion Mining

Sentiment analysis and opinion mining play a central role in Enterprise Text Analytics since they form the back-bone of customer feedback analysis. They also help to understand and measure overall customer satisfaction and identify the different aspects of a product or service deemed important by customers. The overall tonality associated with each object, and the aggregates on each aspect help an organization to identify actionable intelligence items. For example, if a large number of customers express negativity about the rewards program, it is definitely a cue for the bank to relook at it.

Free Text Analysis

Free text (available in the Comments, Additional Inputs, Suggestions, and other fields of various feedback forms) analysis can unearth aspects that are completely unexpected or unanticipated. For example, while analyzing customer feedback about a newly launched credit card for a renowned bank, it was observed that a significant number of customers stated that the “card did not reach them on time” and hence was not accepted or used. This was interesting since this was neither a feature of the card itself nor was it something directly related to the bank’s activity. Further, these customers mostly gave a poor overall numerical rating to the bank, though they were satisfied with the features of the card itself. The overall Customer Satisfaction Index of the region improved considerably by replacing the courier company. This example, and many others, revealed that useful information often comes from the free-text column and not always from answers to structured questions in a survey.

Context Based Analysis

Often, interpretation of text in isolation can be quite tricky and may not even lead to the correct conclusion. For example, while analyzing free text user feedback for an online service provider, a statement that was encountered very often was “I can't find the items that I want on your web-site.” This could simply be interpreted as ‘content missing’, but when feedback from multiple users was analyzed the interpretation that appeared to be statistically more substantiated was that “the search mechanism of the web-site was poor.” The content management team concluded that this was correct.

Statistical Text Analysis

Text is generated in an uncontrolled fashion and is often grammatically incorrect, contains spelling errors and may even contain non-textual elements such as icons, characters, symbols and so on, which may or may not be contextual. Text that is obtained from blogs or discussion forums can be topically noisy, i.e. not relevant to the context of the discussion. In such circumstances, statistical analysis of unstructured text is adopted because commonly used Natural Language Processing (NLP) tools such as Part-of-Speech (POS) tagger, Parser and Named Entity Recognizer do not always operate correctly on noisy text. Statistical analysis involves deriving information through patterns and trends through statistical techniques such as classification.
Banks and Big Data Analytics

Driven by the increasingly competitive landscape, banks have been forerunners at adopting new technologies. According to Michael Hickins, The Wall Street Journal [2], quite a few banks have already come up with fairly mature strategies for the use of Big Data. For instance, four large American banks are using huge amounts of data, including social media posts and emails, along with in-house transaction data and publicly available economic statistics from the U.S. government to gain insights into current and emerging consumer trends.

Big Data analytics promises to improve fraud detection capabilities of banks by allowing them to aggregate and analyze all the available information about a customer from different divisions - information about checking accounts, mortgages and wealth management - and additionally fuse this with social media data to gauge the propensity to do fraud. Behavioral attributes of a customer obtained from social media comprise one of the parameters used to assess credit-worthiness. A straight example is – if the social network of a customer has ‘good’ borrowers then the likelihood of him or her also being a ‘good’ borrower are high. This can improve credit-rating methods and also help improve customer service.

Churn prediction is another area that can employ unstructured data analytics techniques. To give an example – a major American bank was plagued with the problem of their customers defecting to smaller banks. By analyzing customer behavior from their website, this bank was able to identify that customers found its end-to-end cash management portal too rigid. It also realized that customers wanted to access ancillary cash management services because of which they were migrating to other financial service firms and used this knowledge to design more flexible online products.

Conclusion

About 95% of data generated everyday all over the world is unstructured in nature and captures large volumes of information. Effective use of this information can help better explain the past, understand the present and predict the future. Language, expressions and communication mechanisms are constantly evolving. While it would not be presumptuous to state that none of the systems will ever be perfect, with Big Data technologies, banks can leverage the power of unstructured text data to expand their business. However, gleaning meaningful insights from unstructured text is an endeavor where humans will have to remain in the loop to exploit the power of machines which can aggregate, count and report but not think!

About the Authors

Lipika Dey
Dr. Lipika Dey is a Senior Consultant and Principal Scientist at Tata Consultancy Services where she heads the Web Intelligence and Text Mining research group at TCS’ Innovation Labs. Lipika’s research interests are in the areas of content analytics from social media, social network analytics, predictive modeling, sentiment analysis and opinion mining, and semantic search of enterprise content. She focuses on seamless integration of social intelligence and business intelligence. Prior to joining the industry in 2007, she was a faculty member in the Department of Mathematics at Indian Institute of Technology (IIT), Delhi. Her work has been published in several International journals and she has refereed conference proceedings. Lipika has a Ph.D. in Computer Science and Engineering from IIT, Kharagpur.

Sandeep Saxena
Sandeep Saxena leads the solutioning initiative from the CTO Evangelize team and is responsible for the proliferation and application of research work from across the CTO innovation labs, internally within TCS and with TCS customers. As part of this process, his key responsibility involves conceptualizing solutions for a specific domain founded on the research assets from the CTO labs. Sandeep also heads the GTM strategy for TCS’ COIN partners’ solutions. Working with customers from different industries, Sandeep has gained a good understanding of processes across banking, financial services, insurance, retail, CPG, shipping and healthcare domains.
Spandeep is a B.E. (Hons.) in Electrical and Electronics and an M.Sc. (Hons.) in Economics from BITS, Pilani and has been with TCS since 1995.

Chandershekher Joshi
Chandershekher Joshi is an Innovation Evangelist for the Big Data, Analytics and Web Intelligence research areas of TCS Innovation lab. He is into strategizing, consulting, designing go to market plans and solutioning on TCS research offerings.
Chandershekher is a B.Tech and a management graduate from XLRI with more than 7 years of industry experience in the retail, healthcare insurance and hi-tech domains.

References
**About TCS' Banking and Finance Solutions**

Over the past four decades, TCS has partnered with multiple clients in the BFS world and has executed a number of complex and time critical assignments under challenging business and operating environments. The BFS world sees TCS as an organization with strong foundation and superior understanding of the Financial Services market. This paved the way for creation of unparalleled vertical expertise.

Our end to end offerings, comprehensive product suite, scalable processes and innovative frameworks have enabled significant strategic value creation for our clients by helping them optimize their IT investments, enhancing operational efficiencies, minimizing risk, and acquire sustained cost leadership.

In the BFS industry, TCS is ranked high, at No. 2 in the FinTech 100 ranking, and has been among the top 10 companies for the last five years. With 12 out of top 20 global FIs as customers, the clientele include big and small banks, development institutions, regulatory institutions and diversified & specialty finance institutions.

The BFS ISU works closely with Financial Services institutions across the Think, Build & Operate space to fulfill their strategic & tactical objectives. The sub practices include Retail banking, commercial / corporate banking, Capital Markets (Investment, Wealth Management and Securities), Market Infrastructure, Cards (Credit, Debit, Loyalty), Risk Management and Treasury. The TCS risk management framework has been rated as a leader by the financial research firm, Tower Group. Everest Research has also ranked TCS as the leader in the banking application outsourcing space.

**Contact**

For more information about TCS' Banking & Financial Services, contact us at bfs.marketing@tcs.com

---

**About Tata Consultancy Services Ltd (TCS)**

Tata Consultancy Services is an IT services, consulting and business solutions organization that delivers real results to global business, ensuring a level of certainty no other firm can match. TCS offers a consulting-led, integrated portfolio of IT and IT-enabled infrastructure, engineering and assurance services. This is delivered through its unique Global Network Delivery Model™, recognized as the benchmark of excellence in software development. A part of the Tata Group, India’s largest industrial conglomerate, TCS has a global footprint and is listed on the National Stock Exchange and Bombay Stock Exchange in India.

For more information, visit us at www.tcs.com

---

**IT Services**

**Business Solutions**

**Consulting**