Overview

Up until now, despite possessing a large stash of structured data sources in their databases, banks found it virtually impossible to glean 360-degree views of their customers for intelligent decision making. The recent emergence of technologies designed to address 'big data', however, has ensured that today, your bank can have the capability to process huge volumes of unstructured data, typically in the form of e-mails, SMS text messages, customer feedback, etc. To remain competitive in a cut-throat business environment and create tangible value for your customers, your bank requires an integrated data analytics solution that can process and analyze the deluge of unstructured data generated every day to make effective, customer-focused decisions.
Structured vs. unstructured data

As per data scientists, out of the total data that is generated, 80% and upwards is constituted by unstructured data vis-à-vis 20% or less from structured data.

Constituents of unstructured data

What can your bank do with unstructured data?

To differentiate your bank from those in the competition, it is imperative to make use of the large pool of available data to offer targeted products and services designed to cater to your customers’ specific needs. The advent of technology-enabled social media platforms (Facebook, Twitter, Google+, et al) and their almost addictive popularity among millions of users is resulting in the generation of increasingly large volumes of unstructured data on the Internet and mobiles.

The question is this: Is there a way your bank can combine this unstructured data with the structured data in your databases to gain incisive insights into your customers' behavior and preferences for improved decision making?
With the advent of Hadoop, MapReduce, Web 2.0, and related technologies, your bank can now leverage its processing and storage capabilities by complementing it with natural language processing (NLP) and semantics analytics. This can help you more accurately capture your customers’ needs and preferences.

In this paper, we’ll take a look at how effective analysis of unstructured data can help you increase customer satisfaction and loyalty by designing better products and targeting them at the right audience.

**Why should your bank analyze unstructured data?**

*Check this out for starters: Sample of real customer complaints on the Internet:*

**Complaint 1**

“Lousy Customer Service
I want to know about the redressal mechanism that I have as a citizen against lousy customer service by XXX Bank. I am tired of dealing with the same XXX branch of the bank for so many years. What redressal do I have against lousy customer service in the branch? What redressal do I have against incompetent branch staff?
They don’t reply to mails. They make all false promise. They are a bunch of liars.
When I go to their site I find nothing on customer service. So where is an ordinary citizen supposed to go?
Please help me get to some senior folks in XXX bank to complaint against the XXX branch.
Suffering customer!
Joseph”

**Complaint 2**

“I am very old customer of XXX bank. XXX bank’s Internet banking is best in class and it is far ahead than its competitors. Despite of it, other banks are providing statements online, there is no need for the customer to visit the branch for the statement.
Recently, it happened twice in last 1.5 months. I had gone to XXX Branch and XXX branch for statement. During both occasions systems were shut and employees of the bank have no idea of when it will be up. There was utter chaos in the branch.
Bank needs to seriously look into this problem otherwise you will lose out in the competition. There are many private banks in the race; your bank should not lose customers because of downtime.”

The core objectives of every bank are gaining—and retaining—customer loyalty, profitability, and a solid brand reputation. This can be achieved by offering distinct value to customers in the form of differentiated services and products tailored to suit their needs. Traditionally, financial institutions adopted a bank-centric approach that focused on selling products they ‘thought’ their customers wanted but today, a paradigm shift in approach means that banks are required to be customer-centric, offering solutions the customer actually needs.
What insights can a bank draw from the above complaints?

- An indication of the customer’s mood (based on his tone)
- Level of customer service in the branch in question
- What is the customer not happy about?
- Feedback regarding the bank’s web portal
- Feedback regarding the bank’s staff of that particular branch
- Inconvenience caused due to system downtime

Corrective measures that can be taken based on the above insights

- Improve customer service in the branch
- Understand what the customer expecting
- Improve the portal design
- Ensure proper training for employees on customer service

Customer expectations have changed as a byproduct of increased competition. Today’s customer expects your bank to understand his needs and offer him a product that aligns to his specific requirements. The emergence of social media has played a significant role in bringing about change. While social media has encouraged individual customers to have bigger, more unique expectations than ever before, it also gives your institution the ideal platform to identify and understand these expectations.

To add to the complexity, banking is witnessing an increasing shift from ‘brick and mortar’ branches to computers and mobile phones, with modes of payment moving from cash to electronic transfers. While greatly enhancing customer convenience, this has also had the undesired effect of increasing opportunities for fraud. By using cutting-edge tools to pull and analyze data from structured and unstructured sources, it is possible to create a pattern to trace fraud. According to a recent PC Advisor report, banks have witnessed a dramatic decline in certain types of fraud by utilizing valuable insights present in customer data. In 2011, losses from one type of payment card fell to US$321 million from US$515 million in 2008.

According to data scientists, most of the data that is generated (85% and upwards) in banks is unstructured data. This includes e-mails, documents, customer feedback, customer complaints, Internet blogs, and websites. This data is mainly composed of text rather than numbers. However, the effort made to utilize this unstructured data to create significant customer intelligence has been minimal.

Unlocking the potential of unstructured data is crucial if your bank wants to differentiate itself from the competition and be a market leader.

Technology challenges for processing unstructured data

- Storage and processing capabilities to handle huge amounts of data
- Understanding unstructured data
- Converting unstructured data into relational data
Reference architecture of an analytical solution designed to process unstructured data

The solution leverages distributed technologies like Map Reduce (Google's) or Hadoop (open source) to process huge volumes of unstructured data, eventually this data forms input to the analytical engine which generates output for decision making.

Considerations

- Assess data readiness to determine reliability of sources.
- Modify enterprise data warehouse schema to accommodate unstructured data
- Implement distributed file systems while ensuring consistency of foreign keys
- Apply algorithms and stochastic models to the data for processing and determining patterns.
- Populate consumer insights using reporting /visualization tools like Business Objects, Cognos, etc.

How can your bank convert unstructured data into meaningful data?

The conversion of unstructured data into meaningful data forms the crux of the solution. The ‘raw’ text, which is typically in the form of tweets, Facebook postings and comments and emails, is analyzed by parsing the text into words or phrases. These words or phrases are flagged as ‘good’ or ‘bad’ to analyze the sentiment of the customer. For example the word “good” gets a “+1” and the word “bad” gets “-1” and “neutral” get “0”, Subsequently, all the unstructured data in the form of text is converted into structured numeric extracts, which make up the input for analytical algorithms.

Most analytics-based solution for converting unstructured data into meaningful customer insights would typically involve technology elements like text mining, natural language processing, and stochastic-based algorithms.
Natural language processing

NLP has been a prominent research topic for many years now. It is the process of extracting meaningful information from natural language input and producing natural language output. NLP identifies the text (entity) and its characteristics (attributes), which exist in descriptive forms, in compliance with grammar. NLP takes the text as input and generates output relating to the text.

Semantic processing and ontology repository

Although NLP can help in generating entities and their characteristics, it does not understand associated semantics. During semantic processing, such entities and attributes are transformed so they can be translated to the context of target domain. An ontology repository is the hierarchical structuring of knowledge by segregating them according to their essential qualities. Semantic processing uses this ontology repository for transforming entities and attributes into meaningful customer insights.

What are the potential challenges that continue to exist in this area?

- **Accuracy**
  
  There is always scope for improvement; banks have to strive for more accuracy

- **Gap in demand and supply**
  
  Lack of skills is the main inhibitor; most people are familiar with spreadsheets and relational databases but less familiar with unstructured data tools

- **Data becomes obsolete quickly**
  
  As a huge quantity of social media consumer content, including millions of Twitter tweets and Facebook comments, gets generated every day, the shelf-life of data is quickly becoming very short.
Conclusion

The problem your bank may face in analyzing unstructured data to gain insights into customer behavior for improved decision making can be addressed with two-pronged approach:

- Leverage technologies like Hadoop Map Reduce to handle huge volumes of unstructured data that is generated from different data sources
- Use NLP and semantics analytics-related technology to mine text and find patterns in the context of the target domain

About the Authors

Sowmya Ramachandran
Sowmya has over 10 years of experience in the domain of consumer banking. Before joining Infosys, she was part of a leading IT company for 5 years. She brings with her a blend of domain and technology expertise in key areas of banking such as cards, mortgages, consumer marketing and risk and compliance. She can be reached at Sowmya_r08@infosys.com.

Kalyan Malladi
Kalyan has over 6 years of experience in the domain of consumer banking. Before joining Infosys, he had stints with one of India’s top private banks and a multinational bank. His areas of interest include applying analytics to facilitate cross-sell/up-sell opportunities in banks and identification of customer requirements to enable need-based selling in banks. He can be reached at kalyan_malladi@infosys.com.

REFERENCES

Booz co web social media analytics
Semantic driven customer insights by Carey Chou and Kishor Gummaraju
About Infosys

Infosys partners with global enterprises to drive their innovation-led growth. That’s why Forbes ranked Infosys 19 among the top 100 most innovative companies. As a leading provider of next-generation consulting, technology and outsourcing solutions, Infosys helps clients in more than 30 countries realize their goals. Visit www.infosys.com and see how Infosys (NYSE: INFY), with its 150,000+ people, is Building Tomorrow’s Enterprise today.

For more information, contact askus@infosys.com www.infosys.com