MEDALLIA INSTITUTE

Discovering business insights from the voice of the customer.

Medallia's Approach to Text Analytics

by Ji Fang, Ph.D.

TABLE OF CONTENTS

Discovering business insights from the voice of the customer.

The last decade has seen an explosion in the variety of ways customers provide feedback to businesses. More than ever, customers want to make their voices heard inside corporate walls. And they now have a multitude of communication channels available to them: Twitter, Facebook, TripAdvisor, surveys, and more.

This deluge of Voice of the Customer (VOC) feedback includes both quantitative and qualitative information. The quantitative information (e.g, numerical scores) measures overall customer perception, while qualitative information (e.g, free-form text comments) provides the underlying reasons for those overall scores.

Quantitative data is structured and easy to process using traditional business intelligence solutions. Unstructured free-form text data, however, is often stored but not used at an aggregate level for further insights. Given the massive increase in qualitative feedback data from so many different channels, text analytics technologies need to take this wealth of unstructured data and make it actionable. These solutions must be able to quantify the qualitative to highlight areas that have the biggest impact on customer loyalty for specific parts of an organization.

> "Text analytics solutions must be able to quantify the qualitative to highlight areas that have the biggest impact on customer loyalty for specific parts of an organization."

A good text analytics solution

1. Uncovers the reasons behind the scores

Feedback scores tell you what happened; comments tell you why. Text analytics helps you uncover the reasons for changes in numerical scores and identify the locations, employees, and business units responsible for those experiences.

Suppose you are the GM of a large hotel, and you notice that your scores for breakfast have dropped dramatically. Text analytics helps you take action on this feedback by analyzing all breakfast-related comments without having to go through them one by one. With text analytics, you may discover that the problem was powdered eggs, or the canned fruit you had to stock due to a fresh fruit shortage.

2. Analyzes the entire customer experience journey

Businesses design traditional surveys with hypotheses about the customer's experience, and ask numerical questions based on specific experience attributes. It's often only in the open-ended comments that customers can talk about their experience as a whole. Text analytics processes these open-ended questions to uncover the entire customer experience journey.

For example, a call center might ask its customers some score-based questions about the professionalism or helpfulness of its agents. Beyond these specific questions, the customer can use the open-ended comment question to explain that the agent was very professional, but the experience was terrible because he was put on hold for 20 minutes before even reaching the agent in the first place. Text analytics uncovers insights like this without the need for a specific question.

3. Lets you shorten the length of your survey

The world of VOC feedback began in market research departments that sent hundred-question surveys to a select group of customers. Many companies still send extremely long surveys to get feedback on very detailed aspects of their business. Text analytics lets you shorten the length of your survey while retaining the same level of insight, by categorizing comments into areas that correspond to your business processes.

4. Identifies emerging trends

Customers talk about what matters to them, when it matters. Capturing these trends as they come up is extremely important, but specific survey questions can't anticipate all the topics that may arise. An eagerly awaited product might be out of stock, or a price increase might be viewed as unfair. Text analytics allows you to uncover countless issues, opinions, and even opportunities, so you'll know about emerging trends before they balloon into giant problems and missed opportunities that are difficult to resolve.

> "It's often only in the open-ended comments that customers can talk about their experience as a whole. Text analytics processes these open-ended questions to uncover the entire customer experience journey."

5. Allows you to allocate resources more effectively

Many customer-centric companies allocate a team of employees to read and manually categorize VOC feedback to drive business decision-making.

Technology has opened the floodgates of VOC feedback. The massive volume of feedback now available makes manual comment categorization unrealistic. Moreover, human perception is subject to change, making the consistency and accuracy of manual categorization even more challenging. Text analytics can automatically, consistently, and accurately do this work in real time.

Text analytics solutions must be able to analyze large volumes of unstructured qualitative data and present them in such a way that they help the company take concrete actions.

The science behind text analytics

In this section we discuss the two main types of analysis that are performed on unstructured free-form text data, namely topic and sentiment analysis. In order to evaluate the accuracy of different text analytics methods, we use the concepts of precision and recall, which measure preciseness and completeness, respectively.

Precision is the proportion of comments that were correctly categorized into a given topic. For example, if a topic analysis system identifies 100 references to the topic "staff attitude" and 90 of the identifications are correct, the system precision for this topic is 90%.

Recall measures the completeness of your topic analysis system. If there are actually 120 true references to the topic "staff attitude," then the system recall for this topic is 75% (90/120).

The nature of Voice of the Customer (VOC) data

Diversity of business-relevant topics

Customer comments tend to cover a broad variety of business topics. For example, the sentence, "Hotel is nicely situated, polite concierge and clean rooms yet I departed feeling a little bit bitter," contains as many as three topics: hotel location, staff attitude, and room quality.

Each customer comment can cover such a diverse set of topics because it represents a longitudinal description of a customer experience journey, which typically touches many different business operations. According to our survey data, most comments with an average length of 30 words include at least two business-relevant topics.

Richness of strong sentiments

When people respond to surveys or post reviews, they tend to feel strongly about what they're writing. Customers normally don't invest time and energy to describe neutral experiences. This means that customer comments are rich in polarized sentiments.

In a Medallia study examining over 80,000 randomly selected sentences from customer feedback, we found that 80% of individual sentences contained at least one strong sentiment. The prevalence of strong sentiment allows you to sort key topics by how positive or negative they are.

1. Topic analysis

Topic analysis categorizes customer comments into business-relevant topics. There are two approaches to categorize comments: rule-based and machine learning.

Rule-based approach

Topic analysis typically uses a rule-based approach. This approach relies on a set of rules to identify the topics under which the system categorizes a comment. We can define a topic rule by the appearance and/or co-occurrence of certain keywords or word pairs that are grammatically related. For example, the following rules can belong to the set of rules for defining the topic of "Staff Knowledge":

i. knowledge

ii. know NEAR where within 1 words

iii. know CONNECTED TO products

Rule i categorizes any piece of text containing the word "knowledge" into the topic "Staff Knowledge". **Rule ii** tags any piece of text in which "know" and "where" co-occur in the order indicated and are separated by, at most, one word. **Rule iii** captures instances in which the words "know" and "product" are grammatically related. Some examples for rule iii include "...knows the products," "...knows her product," and "...knew a lot about the product." A linguistic dependency parser is normally used to identify such grammatically related word pairs.

Linguists, analysts, and engineers normally build rule-based systems manually. When used effectively, the rule-based approach can achieve precision and recall close to 90% and 60%, respectively. Since the rules are built manually, the data set used is typically a subset of the entire set of feedback data. Natural language, however, is tremendously rich, and people express the same reactions in many different ways. A small data snapshot is unlikely to capture the full scope of natural language, which means that manual rule creation can lead to low recall. In addition, implementation cost is high because the manual process is labor-intensive.

To minimize cost, some solutions bypass the initial topic setup process. They extract and display all the fine-grained events mentioned in the data, where events are usually defined by grammatical or semantic relationships discovered by a linguistic dependency parser. The screenshot to the right illustrates the top events extracted using this method for a retail store.

	# Surveys
better↔customer_service	4064
make↔purchase	3801
go↔have_to	3338
love↔person_name	10814
card↔gift	1473
item↔purchase	2649
person_name↔recommend	4790
department↔shoe	7265
more↔sale	3596
go↔store	3645
nice↔very	8158
have_to↔wait	2748
go↔when	2690
find↔hard	2828
help↔need	3899
like↔see	4053
department↔woman	2963
clothing↔selection	4768

This process, however, captures events that may not relate to business-relevant topics. The events also do not map neatly to an organization's processes, thereby producing some noise, as well as data that isn't readily actionable. If you want to take this approach beyond general monitoring and listening purposes, you will have to create a large number of rules to organize all of these fine-grained events and map them to relevant business processes. Ultimately, this effort is similar in scope to the manual rule-based approach described above.

Machine learning techniques

Alternatively, you can perform topic analysis using machine learning techniques such as clustering and classification to automatically discover and categorize topics.

For clustering, a learning algorithm automatically groups similar data points together. Clustering doesn't require a predefined category set, but it does require a way to judge the similarity of data points, which remains a big challenge for conventional clustering approaches.

Classification automatically categorizes data points based on predefined parameters. Two common approaches are supervised and semi-supervised classification. The former yields better accuracy; the latter requires fewer human inputs. Supervised classification relies on a training data set in which every data point is manually assigned a category to train the classifier. The trained classifier can then be used to categorize new data points. In general, the bigger the training data set is, the higher the accuracy the classifier can achieve. However, for some applications, it is not practical to build large manually annotated training data sets. Semi-supervised learning becomes a very appealing alternative in such cases. It normally starts with a limited number of manually annotated data points. It then does one of two things: leverages a clustering algorithm to classify the data points into the category whose manually annotated data points are similar or leverages techniques such as co-training and active learning to automatically increase the size of the training data set, which will be used to train an improved final classifier.

Although these techniques have received a lot of academic attention, they have not been commonly adopted for analyzing VOC feedback in a business environment.

Conventional clustering techniques are successful at identifying coarse-grained news topics organized around named entities, such as names and other proper nouns, which appear at a high frequency in news articles. The coarse-grained similarity of the named entities can serve as a good similarity measurement for the rest of the content. Customer comments, however, often express topics at the sentence level and are thus typically more fine-grained than news topics. There are also many more ways to express the same topic in customer feedback than in the relatively limited language used to express news topics. The richness of language in customer feedback therefore has prevented conventional clustering techniques from achieving high topic categorization accuracy. However, the recent advancement in machine learning, particularly the evolution of deep learning techniques, has led to significant improvement in addressing this challenge. Deep learning has enabled a way to efficiently analyze very large amounts of raw data and uncover the similarity among data points.

> "The richness of language in customer feedback therefore prevents clustering techniques from achieving high topic identification accuracy."

Classification techniques can yield similar or even higher precision and recall than rule-based approaches. However, while supervised learning yields high accuracy, it normally relies on a large amount of high-quality annotated data, which is often difficult and expensive to obtain. The accuracy of semi-supervised learning relies on the effectiveness of the clustering, co-training and active learning techniques, which remain technically challenging.

Medallia's approach

Medallia is continuously striving to maximize completeness and precision. To pursue high-precision topic identification, we adopt the rule-based approach because it yields the best results, as explained above.

We have also developed a machine learning-based topic discovery and categorization system that automatically analyzes large amounts of data, discovers conceptually related data points and groups them together. This approach leverages a unique form of unsupervised clustering, which is complementary to the rule-based approach. It increases the recall of our topic tagging by eliminating blind spots that can arise during the rule building process and augmenting manually identified topics. This automation overcomes the human limitation in handling the richness of natural language.

Regardless of technique, to ensure that our identified topics are business-relevant and actionable, our analysts typically take our clients through a consultative exercise that maps the topics to business processes and customer journeys.

2. Sentiment analysis

A fundamental goal for analyzing customer comments is to discover the positive and negative drivers for business. Topic analysis identifies the drivers while sentiment analysis illustrates how positive or negative the drivers are.

There are two competing approaches to sentiment analysis: dictionary/rule-based and supervised machine learning. The table below summarizes their pros and cons.

	Strength	Weakness
Dictionary/ Rule-Based	Generic, Industry-Independent	Low Accuracy: Low Precision, Low Recal
Machine Learning	High Accuracy: High Precision, High Recall	Industry-Dependent

Dictionary/rule-based

The dictionary/rule-based approach relies on a predefined sentiment dictionary consisting of positive and negative words with assigned optional sentiment scores based on the dictionary builder's intuition.

This approach is easy to implement and does not require industry-specific tuning, because it assumes that human sentiment can simply be captured by a list of explicit opinion words whose meanings do not change in different contexts. However, it is precisely because of this assumption that this approach often yields poor accuracy.

First, sentiments are often expressed implicitly, without explicit opinion words. For example, the sentence below expresses a negative sentiment about MyTel without any explicit negative words. A dictionary/rule-based approach will most likely analyze this sentence as sentiment neutral.

Paying off my phone bill so I can leave MyTel Mobile.

Second, sentiment, by nature, is context-specific. The meaning and sentiment of a phrase varies depending on the domain or topic. For example, while both sentences below share the same sentiment word—"long"—it yields a positive reading in one case and a negative reading in the other.

The battery life is long.
The check-in line is long.

Furthermore, sentiment reading can change in different question contexts, as shown below.

Q: What compliments do you have for our staff?
A: Housekeeping.

Q: What could we have done to improve?
A: Housekeeping.

Third, due to the complexity and subtlety of natural language, even seemingly unambiguous opinion words often do not yield the correct sentiment reading. For example, the sentence below expresses a strong negative sentiment even though it contains an explicit positive word, "nice."

"I think it is awfully nice how mytel mobile wont allow u to receive calls or take calls when u don't pay yo bill."

For these reasons, the dictionary-based approach often suffers from low precision and low recall. Our study shows that for the task of determining whether a sentence expresses a positive, negative, neutral, or mixed sentiment, the precision of even state-of-the-art dictionaries ranges from 50% to 70% depending on how well-formed the data is. The recall of this approach often falls under 50%.

Supervised machine learning

By contrast, supervised machine learning techniques have proven highly successful in overcoming all the dictionary-based challenges described above. Supervised machine learning relies on a large set of training examples—sentences annotated with sentiment scores as judged by a human. The learning algorithm can then pick up both explicit and implicit sentiment cues, as long as they are present in the training data.

Different domain/context sentiment models can also capture the domain/context-specific sentiment readings. For example, a hospitality sentiment model will learn that "The check-in line is long" is negative, and an electronics sentiment model will learn that "The battery life is long" is positive. Similarly, a sentiment model that is sensitive to question context can learn that a short sentence such as "Housekeeping" is positive when it answers a question such as "What compliments do you have for our staff?", but is negative when it answers "What could we have done to improve?"

Because of these advantages, well-trained domain-specific sentiment models can achieve high accuracy, with precision and recall both higher than 85%. The weakness of this approach is that building the different industry/domain sentiment models is both costly and technically challenging.

Medallia's approach

At Medallia, we build industry/context-specific sentiment models using supervised machine learning techniques to achieve both high precision and high recall. We classify each sentence into one of the following six sentiment categories: Strongly Positive, Positive, Neutral, Negative, Strongly Negative, and Mixed Opinion. Our precision and recall are consistently above 85%.

It's important to note that no solution can produce 100% precision or recall; and accuracy above 85% is more than sufficient to identify primary themes in customer feedback. Our experience has been that, beyond a certain point, optimizing these metrics don't markedly improve the business insights derived.

Recall that we should expect no more than 20% of the VOC sentences to be neutral (see "The nature of Voice of the Customer [VOC] data," above). Since it is less likely that people are sentiment neutral when they discuss concrete topics, when we evaluate the subset of sentences that discuss business-relevant topics, the percentage of neutral sentences should be significantly lower. While dictionary-based approaches often tag topic-related sentences as sentiment neutral around 40% - 50% of the time, Medallia almost always tags less than 10% of these sentences as neutral.

To summarize, Medallia's solution strives to achieve both high precision and high recall for topic and sentiment analysis. We are currently using a combination of machine learning and rule-based linguistic approaches to ensure that our clients have a complete and precise understanding of their customers' voices.

Delivering actionable insights through Text Analytics

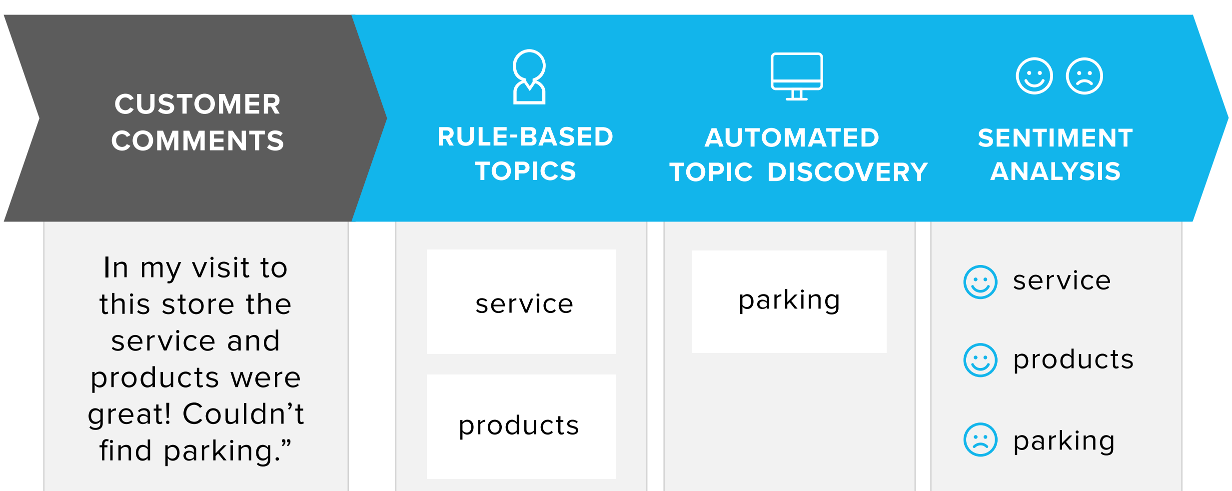
The ultimate goal of any reporting platform is to make text analysis data as actionable as possible for your business.

In light of this goal, Medallia integrates its text analytics solution seamlessly with its core platform, uses relevant scores to measure topic impact, maps qualitative data to specific parts of the organization, and presents this data in real time.

1. Text Analytics is fully integrated within Medallia's CEM platform

Medallia's Text Analytics solution works seamlessly within its core CEM platform. Unlike other solutions in the market that operate separately from the CEM platform, Medallia's Text Analytics is fully integrated to offer consistent and powerful functionality.

THREE ENGINES OF MEDALLIA TEXT ANALYTICS



Medallia integrates scale-based quantitative data with text analytics so that you can easily see the most relevant topics that show strong correlation with low scores.

2. Impact Scores show your most important topics

Medallia calculates the impact that topics have on overall customer loyalty scores such as NPS. The reporting application highlights areas that drive customer loyalty so that you can easily prioritize which aspects of the customer experience to improve.

Proof Point: International Telco

With a survey that features just one scale-based question and one open-ended comment question, the customer experience team at an international telco relies heavily on text analytics data to figure out the current trending topics.

At its quarterly meetings with the executive team, the Medallia customer experience team is able to present trending topics—whether around account activation, call center experience, or phone purchase at a brick-and-mortar location. The team is able to show how much impact these topics have on the overall satisfaction scale-based question, and propose a set of concrete actions to improve the score.

3. Insights map to the entire organization

Medallia's CEM platform maps customer feedback, in real time, to employees responsible for the customer's experience. Employees are empowered to fix processes as soon as issues occur. Medallia is exceptional at representing an organization's hierarchy, which allows users from all branches of the company to locate the data most relevant to them.

Proof Point: Major Multinational Hospitality Chain

Prior to switching to Medallia, a hotel chain was using a text analytics product that wasn't connected to its organizational hierarchy. Because of the complexity of the product and the difficulty in tracking who was responsible for which comments, the customer experience team did not feel comfortable granting product access to anyone outside the team. As a result, turning the unstructured data into something actionable was a long and difficult process.

The hotel switched to Medallia Text Analytics because it was easier to understand, use, and communicate across the organization. The experience team was able to master the platform quickly, and it now provides regular internal reports for all levels of the organization.

4. Real-time insights identify emerging issues and trends

Specialized modules within Text Analytics highlight, in real time, the areas that are having a large negative or positive impact on customer loyalty. Medallia has the flexibility to configure reports showing correlations between customer segments, organizational hierarchy, and comment topics and sentiment, even as customer feedback flows into the application.

Proof Point: Fortune 500 Financial Services Firm

A Fortune 500 Financial Services firm uses Medallia to collect feedback on the customer service at its US call centers. In January, it noticed that cus-tomer satisfaction scores for the Central region had dropped dramatically. The customer experience team was able to use Text Analytics to quickly identify that wait times were particularly bad that month. On further investigation, the team realized that there had been a staff reduction in the Central region due to poor weather, but the calls were not getting routed to other regions for resolution, resulting in longer-than-normal wait times.

Summary

Voice of the Customer feedback contains a wealth of business-relevant topics, conveyed with strong customer sentiment. Text analytics leverages this powerful, unstructured data for business insights and actions. Businesses can use text analytics to uncover the entire customer experience journey, and understand the root causes behind customer feedback. It also helps identify emerging trends, free up resources, and shorten surveys.

Such a solution should have excellent text analysis and reporting capabilities.

- A successful solution automates large-scale text analysis, maximizes preciseness and completeness in categorizing feedback data into business-relevant topics, and accurately tags sentiment.
- The reporting platform should disseminate powerful insights across the organization in real time, enabling specific employees to directly take action on feedback as soon as it's provided. The best solution consolidates analysis from structured and unstructured VOC feedback to point toward the areas with the highest impact on customer experience.

Ji Fang is the lead research scientist for Natural Language Processing and Machine Learning at Medallia. Prior to that, she was a research staff member at Xerox PARC. Ji studied Linguistics at Stanford University and Peking University. Her research in the NLP area has resulted in more than 20 publications published in top peer-reviewed journals and major conference proceedings. Ji also has 5 patents granted and 3 pending U.S. patents.