Customer Feedback Isn't Binary… So why is your survey?

BOULDER EQUITY ANALYTICS

May 2018

"Dark data" are everywhere. By some accounts, upwards of 80% of business data is unstructured, in the form of text, audio, videos and pictures. Great strides have been made in recent years in transforming this dark data into information…taking us into the realm of "big information." The data have always been "big" …just not accessible.

As a key component of this dark data, text is a rich repository of information. Think of all the emails, documents, reports stored on business servers. Machine learning and artificial intelligence (AI) is now making this information accessible and actionable.

One example, the focus of this paper, is customer feedback provided in open-ended survey questions. Traditionally such feedback was "processed" manually by analysts who spent countless hours categorizing feedback into a manageable number of categories or themes and reported out in Excel or PowerPoint.

Now AI technology can automate this process, making it more efficient, consistent and replicable and freeing analysts to do what they were hired to do – analyze.

And making the "gems" in open-ended customer feedback more visible and actionable to management and operational departments so that corrective measures can be made.

Key Takeaways

- **Fragmented text challenges Boolean and NLP logic**
- **AI technology excels at summarizing contextual themes**
- **Feedback processing now takes seconds; can be repeated quickly, and efficiently as new survey responses are received**
- **Data visualization enables insights by making the discovered themes accessible and analyzable by linking back to rest of the survey**

Introduction

It used to be the case that processing open-ended customer feedback questions meant countless hours poring over text, often fragmented, of poor grammar and frankly often written by irate customers, with text laced with ALL CAPS and vulgar language. The analysts' task was to make sense of these comments by assigning them each to one or multiple themes, themes which tried to capture the essence of the customer's comment. Then they would summarize these themes in charts and typically report out in Excel or PowerPoint.

The goal was noble – get relevant customer feedback to management and the correct departments so that corrective measures could be made.

But lags in this manual processing time coupled with human errors in interpretation of what customers were saying, inconsistency in interpretation across both time and analysts and voluminous data (e.g. think recurring tracking and intercept surveys with thousands of responses every month) often meant that the desired corrective measures were never fully realized.

Today, machines, after some human training, are performing this classification task. And performing it much faster, more efficiently and more accurately than could a team of analysts. Moreover, the application of machines extends beyond open-ended survey response questions. Any source of "unstructured" data (emails, transcripts, financial documents, contracts, reports, etc.) can be subjected to the same kind of machine classification process, shedding light on what has previously been "dark" data.

So, we are in a new era of "big information." The data were always "big." Just not easily accessible. A ubiquitous statistic, though unverified, is that 80% of business information is in an unstructured form (think of all the text data in a typical corporate server, or your laptop). Now, AI powered software is making this data accessible and turning it into information, into a knowledgebase, that can support the decision-making process.

Here at BEA, we have applied machine learning, along with text analytics and data visualization in several areas. The case of open-ended survey responses is especially instructive as it highlights how the methodology can directly serve a critical business need. In this case, improving customer service.

Unstructured Data

As opposed to "structured" data…think of a spreadsheet with rows and columns of numbers…text is "unstructured." A paragraph of text, though having its own textual, linguistic structure (nouns, verbs, phrases, semantics, etc.), is not able to be processed in a quantitative way in its raw form. At the purest level, a computer processes 0's and 1's. Raw text does not fit into this numeric format.

Among types of text, open-ended survey responses can be some of the "dirtiest" even in terms of textual structure. Customer feedback can range from a few heartfelt words to a well-thought out response with complete sentences and thoughts. A substantial portion of (most?) customer feedback consists of fragmented sentences. Moreover, grammar, spelling and punctuation are not high on the priority list when consumers are providing feedback.

Typically, survey vendors do not attempt to make sense of open-ended responses and deliver them in their raw, unfiltered, unstructured form, either in documents or spreadsheets, looking something shown in Figure 1.

Imagine having to read thousands of these a month, making decisions as to what theme category(s) each response should be placed into.

It should be noted that "unstructured" data consists of more than just text. Videos, pictures and sound also are not structured in the traditional, rows and columns sense. But advances in machine leaning are allowing for quicker, more efficient and more accurate processing of all unstructured data, including text, which is our focus here.

Figure 1 - Raw open-ended responses from survey vendor.

Tracking time of delivery was incorrect but
To have had someone available to check with the Smith facility directly to see if the driver has returned there, or if the driver is still in Newtown. Some trucks have already returned to Smith. It would be extremely helpful to know if my package arrived back there now. I was told that wasn't possible and to call customer service. The automated system was unable to direct me to a customer service representative. Thank you.
This organization charged me 50 dollars to deliver a package on time, 4 days later no package,
This is what I wanted. Company is completely useless. I wish I would've found a different carrier for my package. Started out I should have got my package Wednesday, then it shows I will get it the 23rd. Now it says the 29th. I will never buy from anyone else that ships with Company.
This is the WORST service that I have ever received.
There's no place to leave a complaint!! Service Express takes another 4 days to get my package and I was told I would have it in 5-7 business days. If Company had delivered it would have been here today!!
There was no place for me to complain on how lousy your shipping service is. This company seems to be unable to deliver package on time. Always 2 to 4 days past delivery date. Pathetic
There was no information regarding my package when it will be delivered. I was told it would be delivered on the 5th. That never happened
There is nowhere on your site that will give me info on why my package has been stopped in Salt Lake City for 5 days. This package is important, and it is already past the delivery date
There is no way to contact someone. My $1500 computer was left on my neighbor's porch with no one home and no signature. I was home right next door
There is no phone number to contact support
There is no explanation of what is wrong with my shipment and no person available to help
The website is unorganized and unnecessarily complicated. The service was abysmal. I much prefer your competitor.
The website asked for my address but did not ask present or new, and it set an account. I have moved and require being notified when packages arrives, I have lost packages many times.
The website as the process is designed the wrong way around, customer need to fit the Company framework rather than the other way around

State of Text Analytics

In the beginning, there was George Boole…

The Laws of Thought, written by George Boole and published in 1854, introduces Boolean logic, which is arguably the foundation of the information age. While many companies still use manual coding and some crowd sourced services, machine processing of unstructured text data started with simple keyword/Boolean search. You still see published "insight" which simply aggregates word counts or uses a "bag of words" technique to approximate themes (i.e. what the text "is about").

Figure 2 - Example of Boolean search string.

Automatic Speech Recognition

(automatic speech recognition) NOT (intravitreal ranibizumab or Invesco Mortgage Capital or intravaginal ring or Bernard Bonnefond or Investigator Resources Limited or Almac Group or vapor gas or Violence Reduction Unit or forensic psychologist or Institute for Virus research or VRU Towers or mortgages or clinical trials or condominiums or IRCTC OR RST=BOARDR OR HD=MFP Suggestions)

This remains by far the most common methodology in use today. It is a good start, certainly more scalable than a team of analysts/coders.

Boolean operators are very precise in that they apply a set of math/logic operations on words. If a keyword(s) is in a bit of text, then it returns "true," what we call a "hit." As you work with the expression, you can eliminate false positive results with exclusions (e.g., if it has "delivered" then "true"; but "false" if it also has "on time" assuming you are looking for delivery problems).

For the analyst, this is a time consuming and iterative process as the Boolean operators need to be finely tuned over time to exclude noise. One analyst told us that to set up 60+ Boolean operators on a news feed took 2-3 months to tune.

Of critical importance to classification of text into themes is Boolean logic's true or false result. There is no "almost" capability.

Boolean operators don't include misspelled words, so what is essentially an exact hit can be missed. For example, "on time" is returned as "true" while "on tme" is evaluated as "false."

Boolean operators are also subject to the analyst's bias, in that the Boolean string is created based on what the analyst thinks is in the open-ended responses. If the word isn't in the string, then that "hit" will be missed completely.

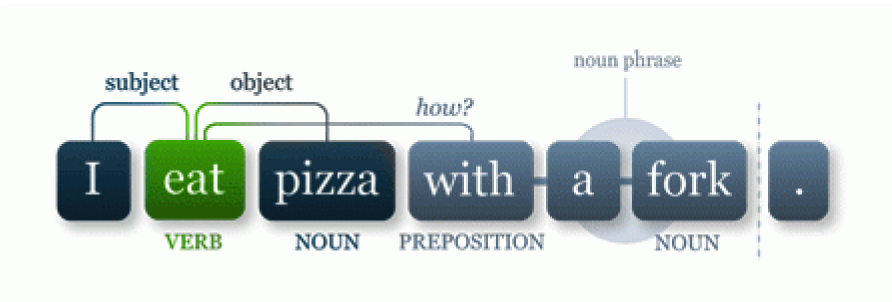
…then came NLP…

Natural language processing (NLP) dates to the 1950's from which rules were developed to help automatically translate text. With the advent of the "statistical revolution" in linguistic analysis in the 1980's and 1990's, NLP became a leading machine learning technique for analyzing text.[1]

By breaking text apart into its constituent parts, NLP algorithms seek to reverse engineer language. Once broken down, the parts can be reassembled based on the desired end.

With large training datasets, NLP has become quite good in many applications. For example, when paired with grammatical rules, NLP can be used in translation algorithims. NLP is also used in search routines (e.g. Google search), spellcheckers, to detect email spam based on the presence of certain words and phrases, by chatbots and resume reviewers.[2]

Figure 3 - NLP sentence breakdown.



NLP is also used to determine the "sentiment" of text. For example, based on the presence of certain positive and negative words and phrases, NLP can be used to ascertain whether a Yelp review has a "positive," "neutral" or "negative" tone. A fairly large body of work has been done in this area with respect to financial documents and earnings call transcripts. Specifically, the academic research has focused on whether executives' "tone" can be tied to future company performance.[3]

However, for classifying a chunk of text in terms of context or theme, NLP suffers from some of the same shortcomings as Boolean logic. It works best on well written or professionally written documents. But extend it to fragmented text, run-on paragraphs, incomplete/poorly structured sentences and the large variety of writing styles/quality found in open-ended survey responses, and it will likely under perform.

But more importantly, while generally good at identifying nouns, verbs, adjectives and the like, and at assembling them into sentences, NLP is not necessarily good at discovering how those sentences (and their components) fit together. That is, NLP struggles to identify the *theme* being expressed.

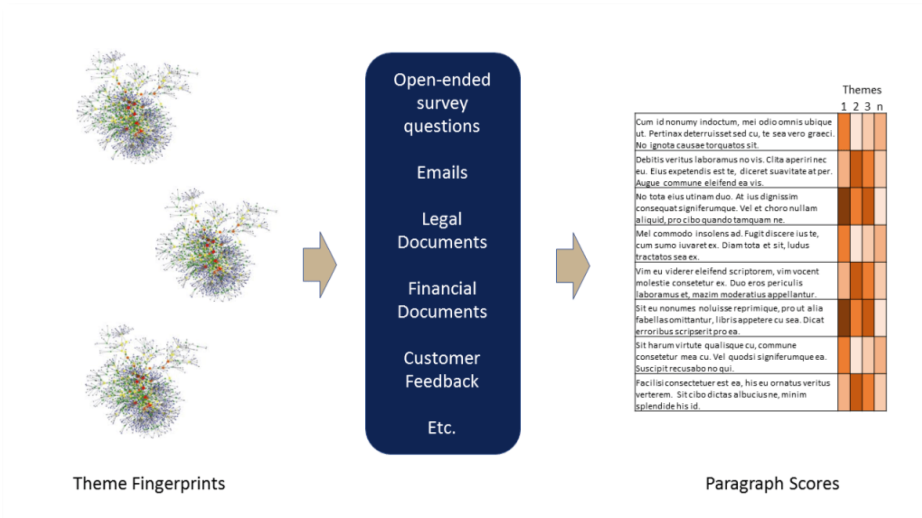
…and then the rise of Artificial Intelligence

Advanced machine learning techniques, often synonymous with artificial intelligence (AI), have taken text analytics to a new level.

[1] See https://en.wikipedia.org/wiki/Natural-language_processing
[2] https://www.cio.com/article/3258837/artificial-intelligence/what-is-natural-language-processing-the-business-benefits-of-nlp-explained.html
[3] See Loughran, T. and B. McDonald, "Textual Analysis in Accounting and Finance: A Survey," Journal of Accounting Research, 54(4), 2016.

Algorithms like that used at BEA, process each paragraph and line of text much the way our brains do, learning the patterns of language, the “key” words, their importance and the words most closely associated with them. The AI is intently focused on understanding *context*.

Figure 4 - AI fingerprinting and text scoring.



The AI provides commands to extract as an array those key words and associations, their direction and values (strengths). The AI is looking to understand arrays of words that represent a theme, regardless of whether the words are a noun, verb, adjective, etc.

This is not a case of turning the computer on and letting it run, however. Like most AI algorithms, the computer needs to be trained in what to search for.

Operationally, the analyst trains a set of, what we call, “agents” based on the themes he/she wishes to uncover. Each agent corresponds to one theme. For example, an online seller will focus on certain themes critical to customer satisfaction such as performance of the online application, product delivery, customer service and the like. Each of these would be a separate agent.

Agents are trained on examples of text (phrases, sentences, even whole paragraphs) reflective of the targeted theme. In addition, keywords are added to the training set to help “boost” the agent’s predictive ability.

Once trained or “fingerprinted”, these agents “read” and score the thousands of open-ended responses for how similar they are to the theme on which the agent was trained. The greater the similarity between an open-ended response and the training theme, the higher the score returned by the agent.

This is more than NLP as the AI is searching for complex word patterns that it has learned from the training dataset, regardless of whether the sentences are whole and well-behaved. Although it is possible for a very well-trained NLP algorithm to be competitive with our AI process, the inherent rule-based nature of NLP renders its less flexible and adaptive, not to mention it works best when the text is clean and well-structured…not what you find in open-ended survey questions.

The scores returned by the agents are numeric, continuous values. The advantage of this approach is that the analyst can use the score, to determine how closely the open-ended responses are matching the targeted theme. By rank ordering the scored responses, the analyst can finely adjust the “cut-off” score used to classify a given open-ended response as a match (yes/no) to the targeted theme.

Importantly, once trained, agents are available for use on multiple surveys (e.g. tracking surveys) or similar feedback in emails or customer reviews. Essentially a **library of trained agents** is created, ready to be deployed across a wide range of text sources.

Overtime, tracking agents’ performance relative to both how they are scoring their assigned themes as well as the number of responses that do not get classified (the “unclassifieds”) informs how frequently agents need to be updated. Typically, we find that agents need to be updated only infrequently

The agents can also be grouped based on functional categories such as product performance, e-commerce website, customer service, shipping, returns, etc. So, if you think of the groupings of agents as clusters of concepts (fingerprints) in your brain, you can apply them in different combinations depending on the task/context in front of you.

In other words, **you can create your own context** and not be forced to adapt to a rigid taxonomic structure.

From Themes to Insights

In practice, clients may or may not have a firm idea of what "themes" they wish to uncover in their data. Having already been through the exercise of manually categorizing text, some clients may already have a starting set of themes in mind. In other cases, the agent development process is much more exploratory in nature.

In both cases, example text and keywords are used to develop a set of agents and are tested against a sample dataset. Through an iterative, collaborative process a final set of agents are derived. These agents are put into production and "deployed" across the entire text corpus.

The result is that every paragraph of text in the corpus (open-ended responses in our example) is scored by every agent. With 40 agents and 5,000 open-ended responses, 200,000 data points are generated.

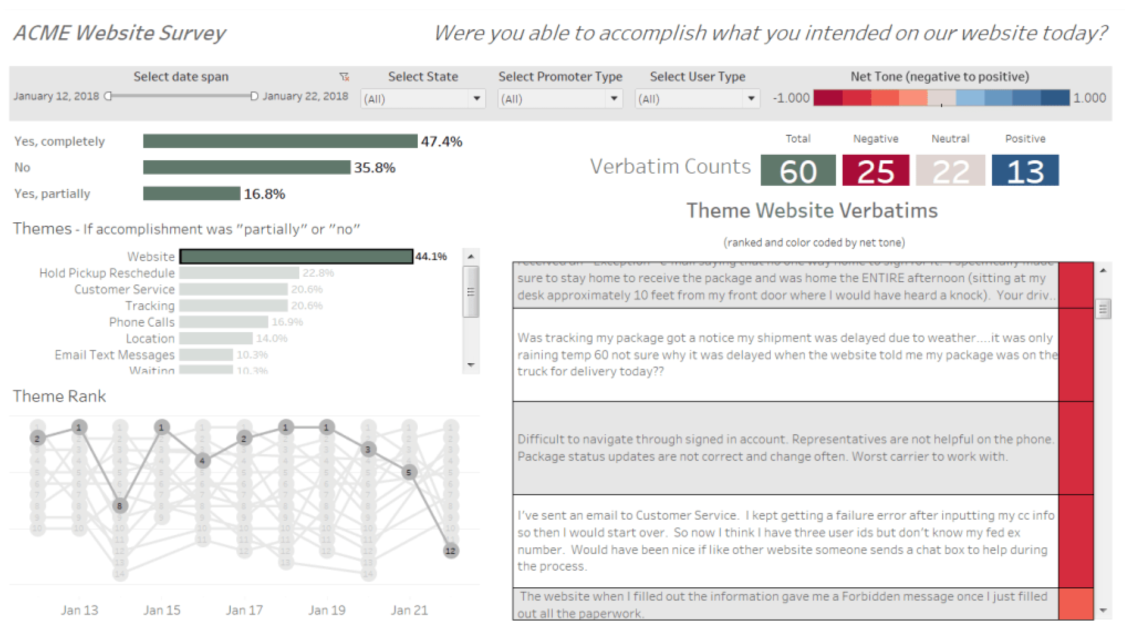
The challenge at this point is how does the analyst make sense of and use this to extract actionable insights?

Our preference is to load the data into a data visualization tool. This enables the analyst to quickly filter, segment, rank and sort the scored, open-ended responses and produce presentation-ready graphics.

An additional benefit from using a data visualization tool is that the analyst can quickly discern any changes when new and or revised data are uploaded. This is especially advantageous when analyzing the results from an ongoing tracking or intercept survey…just like the underlying agents, the data views only need to be built once.

Since the open-ended responses are tied to a survey respondent id, the agent scores can be tied in with the rest of the survey

Figure 5 - BI dashboard: open-ended response themes.



data. This also plays to the strengths of a data visualization tool.

So, rather than building reports in Excel and PowerPoint, the analyst can use a dashboard like this to analyze the open-ended response themes:

The primary advantage of presenting the "themed" open-ended response data in a dashboard is that the analyst can now visually see and track what customers are saying in their own words.

Which themes are the most prevalent? What is the sentiment of the open-ended response associated with each theme? Are they all negative? How many are positive? Neutral?

If the customer satisfaction/feedback/intercept surveys are ongoing, how do these themes track over time?

Filtering on specific themes allows the analyst to see and read the specific open-ended responses that have been classified as belonging to that theme (as shown above).

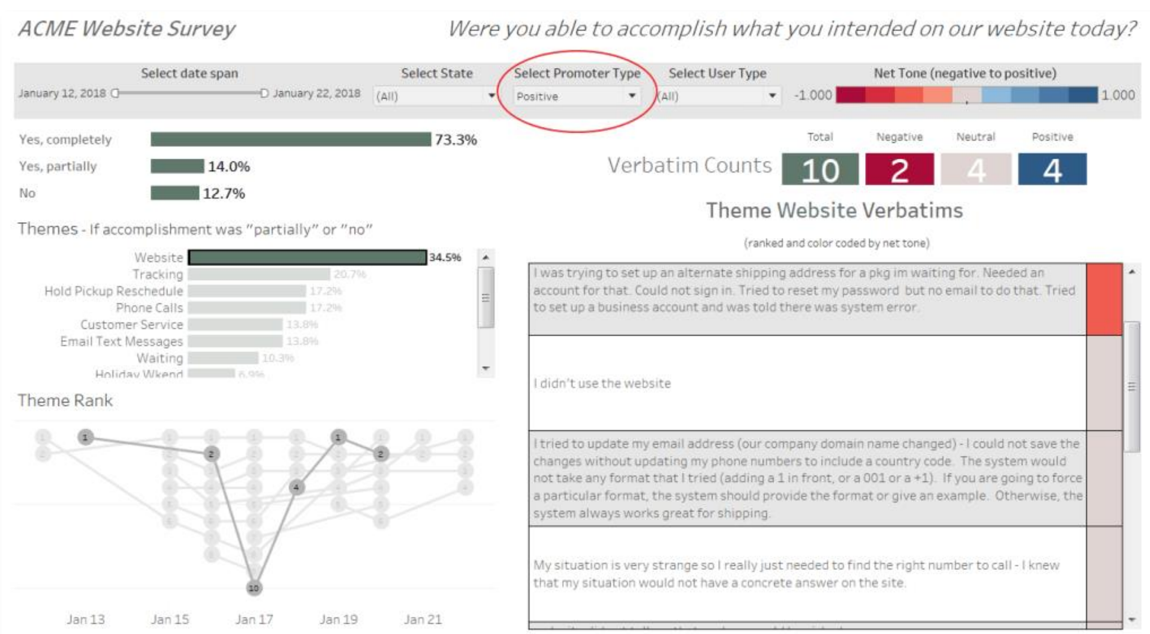
Built correctly, the system should give the analyst full drill down capability from high-level charts to the text of a relevant open-ended response.

Moreover, since the themed open-ended responses have been linked back to the rest of the survey, the analyst can use the rest of the survey questions as filters.

For example, what themes are expressed by "net promoters"? By "net detractors"? Business users vs. non-business users? Do the themes expressed vary geographically (e.g. shipping/delivery issues)?

Rather than spending his or her time assigning responses to themes, importing data to Excel, building charts and then reporting out using PowerPoint, the analyst can spend their time doing what they were primarily hired to do…analyze.

Figure 6 - BI dashboard: filtering based on closed-end survey questions



Accelerating the Call to Action

Segmenting the open-ended responses in this manner now enables the relevant responses to be sent to the appropriate departments for follow-up.

All responses having to do with website performance can be sent to the IT department, for example. Responses having to do with shipping can be peeled off and sent to logistics. Filtering by state or region allows for even finer segmentation and issue follow-up/resolution so reporting can be mapped to the organization.

While some of this could have been done without machine learning technology, the power is in the integration of multiple capabilities in an agile responsive solution…combining the data processing afforded by the machine learning technology with the analytical and reporting functionality afforded by a data visualization tool.

All of this is accomplished in a manner of seconds rather than days or weeks. And the process can be repeated quickly, and efficiently as new survey responses are received.

Conclusion

It wasn't that long ago that analysts had no choice but to toil long hours reading and classifying thousands of open-ended survey responses, pounding Excel and PowerPoint each month to create and distribute reports. With advances in machine learning as well as data visualization tools, analysts have been freed to focus on the insight and the corrective action that benefits their organizations.

Here at BEA, we enable organizations to fully exploit the potential of their unstructured, "dark" data. By deploying our proprietary AI technology against open-ended survey responses and other types of corporate text, we can help you focus on what is required to better serve your customers. For additional information, visit us at www.boulderequityanalytics.com and contact us at eh@boulderequityanalytics.com.