Clarity: Data-Driven Automatic Assessment of Product Competitiveness

Sheema Usmani, Mariana Bernagozzi, Yufeng Huang,
Michelle Morales, Amir Sabet Sarvestani, Biplav Srivastava
IBM Chief Analytics Office,
Armonk, NY, USA 10504

Abstract

Competitive analysis is a critical part of any business. Product managers, sellers, and marketers spend time and resources scouring through an immense amount of online and offline content, aiming to discover what their competitors are doing in the marketplace to understand what type of threat they pose to their business' financial well-being. Currently, this process is time and labor-intensive, slow and costly. This paper presents *Clarity*, a data-driven unsupervised system for assessment of products, which is currently in deployment in the large IT company, IBM. *Clarity* has been running for more than a year and is used by over 1,500 people to perform over 160 competitive analyses involving over 800 products. The system considers multiple factors from a collection of online content: numeric ratings by online users, sentiments of reviews for key product performance dimensions, content volume, and recency of content. The results and explanations of factors leading to the results are visualized in an interactive dashboard that allows users to track their product's performance as well as understand main contributing factors. Its efficacy has been tested in a series of cases across IBM's portfolio which spans software, hardware, and services.

Introduction

Every business wants to know how their product/offering, whether software, hardware or service, is doing in comparison to its competition. Many people are interested in competitive analysis, the primary being marketers, sellers, and product managers. Currently, such users scan through the large volume of online and offline content, aiming to understand what one's competitors are doing in the marketplace for every product they have, to understand what type of threats they may pose to the business' financial well-being. This process is time and labor-intensive, error-prone, slow and costly. Furthermore, as competition and feedback from users continue to evolve, any analysis done previously needs to be frequently updated to ensure accuracy.

To address this business need, we introduce an a deployed system, called *Clarity*, which analyzes the competitive landscape of products in a marketplace continuously, without supervision, as data gets updated over time. We now first preview the working of the system by providing a running example, and then describe its details.

Running Example

Let us consider the competitive landscape for Product-A[1]. We first determine the similar products which Product A competes with. In this example, they are referred to as Products B, C, D and E. The selection of products for a marketplace is a business decision. The output of *Clarity* is visualized in Figure 1.

All the products are compared based on the *Clarity Score*, which is a numerical value summarizing the online reviews and contents. The *Clarity Score* for each product is displayed through a visualization, as shown at the top part of Figure 1.

In the Figure 1, a ranking of the products are plotted to provide a quantitative overview of the competitiveness of the target products with respect to competitors over 12 months. Using different color schemes for different products, the ranking over the predefined time period is displayed. In addition to the ranking, the width for each product represents the normalized *Clarity Score* in that period. As we can see in the chart, Product A has the highest score over the time period considered, thus it was ranked first throughout the chart. However, the ranking could change dramatically across time. For example, Product D (shown in orange) was ranked third at the beginning of the time period, then it was ranked at fourth in the following month, and then the ranking changed again to second. With this information, the stakeholders of the target product can get a sense of how all the players are performing in the market.

To shed light on which factors are contributing to the Product's score, more details about how the *Clarity Score* is calculated for each product is showed in the bottom of Figure 1. For each product, the main contributors to the score are the *number of mentions* and the overall *Sentiment score*. The sentiment score is the aggregated value of the 5 drivers of the product. As we can see, this gives a more granular level of information of how the products are compared. For example, although Product A has an overall higher score than Product B, Product B receives a higher average *Sentiment* score in its price. However, Product B has a much lower *Compatibility* driver score, which is the main contributor to its lower overall score.

Current users of *Clarity* use the score in their workflow

[1]Due to business reasons, product names are anonymized in the paper.

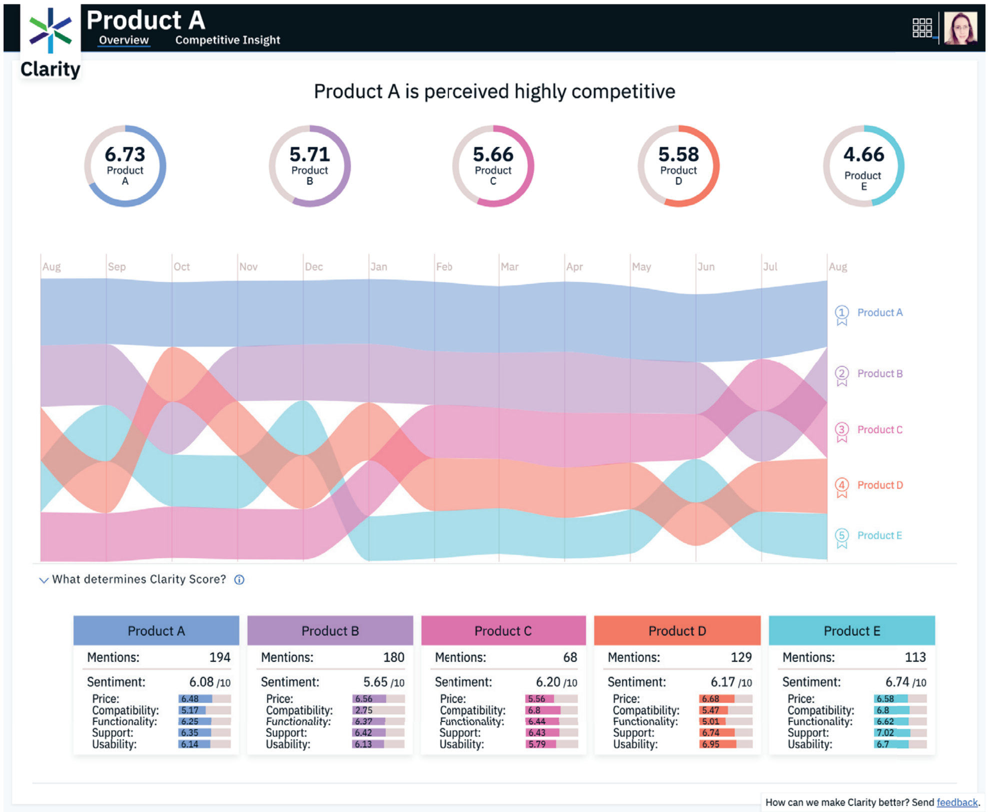

Figure 1: Sorted stream graph to visualize products competitiveness. In the middle section of the graph, X-axis represents time, Y-axis corresponds to product rank and the thickness of line corresponds to absolute competitiveness score.

to understand the competitive stance of their product in the marketplace and leverage the detailed factor analysis to understand their products' strengths/weaknesses as well as those of their competition. Together the high level and detailed level analysis help users make data-driven, informed, decisions regarding the strategic development plan of their products.

Contributions

Business development involves communicating the value of an organization's products to potential customers. *Clarity* helps our business develop by allowing end-users to better understand the value of their products. Our contributions are the following:

- A novel unsupervised approach to assess the competitiveness of products in a marketplace along factors learned from data.
- A novel approach to explain factors affecting competitiveness score of products and visualization of the results.
- Evaluation of the implemented and deployed system, *Clarity*, which has been running for over a year and used by thousands of users assessing hundreds of products. The comparative evaluation with market experts, show that our system is aligned with the competitive analysis by the market experts.
- Integration of *Clarity* into many business applications programmatically that are further accessed by additional hundreds of users daily.

The remainder of the paper is organized as follows: we start with the background and related work, then provide a system overview of the deployed application. Next, we discuss the evaluation of the system and its performance in the field. Finally, we conclude with a discussion and future work.

Background

In this section, we will discuss the competitive analysis process and related effort so that our work and the contribution of our system can be better understood.

Market Intelligence Process

In the field of market research, comparative analysis of different products has been mostly a manual process. The researchers identify top competitive product(s), read through thousands of reviews, keep track of drivers and themes of interest for each product, decide whether a mention represented positive and negative feedback by manually annotating each mention, use the gathered data to make a decision on whether or not product x_1 is more competitive than product x_2, along the drivers and themes considered. Researchers would repeat this process for every additional public domain forum, for every new driver, and continuously revise over time for updates in content.

The manual process outlined above is not scalable especially when one considers the product portfolios of large businesses with hundreds of products.

Related Work

There is a large-scale trend of using computing for business operations like business development (Srivastava et al. 2018), marketing, sales, and product development. Furthermore, Natural Language Processing (NLP) methods, including text mining, are being used to understand many parts of the business landscape including customer needs, product competitiveness, and company performance. Specifically, researchers have surveyed the area of competitive intelligence for products and have demonstrated the promise of approaches using NLP and text mining (Amarouche, Benbrahim, and Kassou 2015).

In Joung et al. (2018), the authors use text mining methods to analyze customer complaints and find gaps in the company's products. In Afful-Dadzie et al. (2014), the authors perform text analysis on user comments posted on social media to compare telecommunication providers in Ghana. In (Bhatt, Mcneil, and Patel 2014), the authors track general sentiment overtime for products by calculating a sentiment score based on user-generated content such as reviews and comments.

Our system builds upon previous work by introducing a novel competitive metric that encompasses sentiment as one of its contributing factors. Our system not only provides a metric but also aims to explain performance, which is a critical step in the market intelligence process. To the best of our knowledge our system presents the first unsupervised approach to ranking and understanding product competitiveness.

System Overview

The main steps of Clarity are:

1. Prepare review data of products p_1 to p_N from sources d_1 to d_M (offline)
2. Process request for analysis for product p_i (online)
3. Visualize analysis results (online, optional)

The steps involved in the computation and visualization of the competitive score for a product are described below. The system can also be invoked programmatically via APIs in which case visualization is not invoked.

Data preparation

For *Clarity* to produce meaningful summaries of online content/reviews, large amounts of text contents are aggregated from an undisclosed set of public forums and review sites, where comments are widely shared by users and experts. We concatenate all the reviews related to a product from a given data source into a single text file. Thereby maintaining a file per product per data source. This standardizes the data storage for systematic downstream processing.

Having the data stored in an unstructured format as text, we can employ various natural language processing techniques to extract information for business insights. Although simple text analytics techniques using keyword extraction or term frequency inverse document frequency (TF-IDF) methods can be used, word vector models are more appropriate for scaling up as more and more data is collected as time goes on. Otherwise, the set of manually selected keywords need to be updated frequently, requiring high maintenance. Moreover, simple keyword models can become too complicated to maintain as the set of words is expanding to analyze new data.

Because *Clarity* is still a relatively new system, instead of using our own Word2Vec model, sets of pre-trained off-the-shelf Word2Vec models are used. Another advantage of using third party models is that it is not necessary to retrain and update the models frequently. The whole process is implemented once and remains the same for consistent outputs. However, as more data is collected, a customized Word2Vec method could be developed. On the other hand, transferred learning models like BERT (Devlin et al. 2018) or GPT (Radford et al. 2018) models can be utilized as well.

Algorithm 1: Data Preparation - Offline

```
Result: Steps to process online reviews
1. documents := retrieve data from source d_1 to d_N ;
2. for document ∈ documents do
    sentences := tokenize the document into sentences,
    foreach sentence ∈ sentences do
        Tokenize the document into words,
        Apply Word2Vec model,
        Save the corresponding word vector
    end
end
```

Analysis

The first component of our system prepares and processes the data used. The system includes an algorithm for data preparation as demonstrated in the high-level description of **Algorithm 1**. As part of the second analysis step, the system generates the *Clarity Score*, which is a numerical value denoting the perceived competitiveness of the product. The score is calculated based on the number of online reviews, the star rating and the aggregated sentiment towards each of the drivers of the product.

The main steps for computing the *Clarity Score* is described in **Algorithm 2**. The calculation of the number of online reviews and the average star rating is done directly by processing the text files generated during data preparation. The calculation of the average sentiment towards each driver includes an NLP engine (outlined in Figure 2), which processes the text in order to understand how the product is performing across any number of drivers/topics.

Using the text as input the engine will extract keywords as well as calculate the targeted sentiment towards those particular keywords, using the Watson Natural Language Understanding API. The trained word vector model is then used to

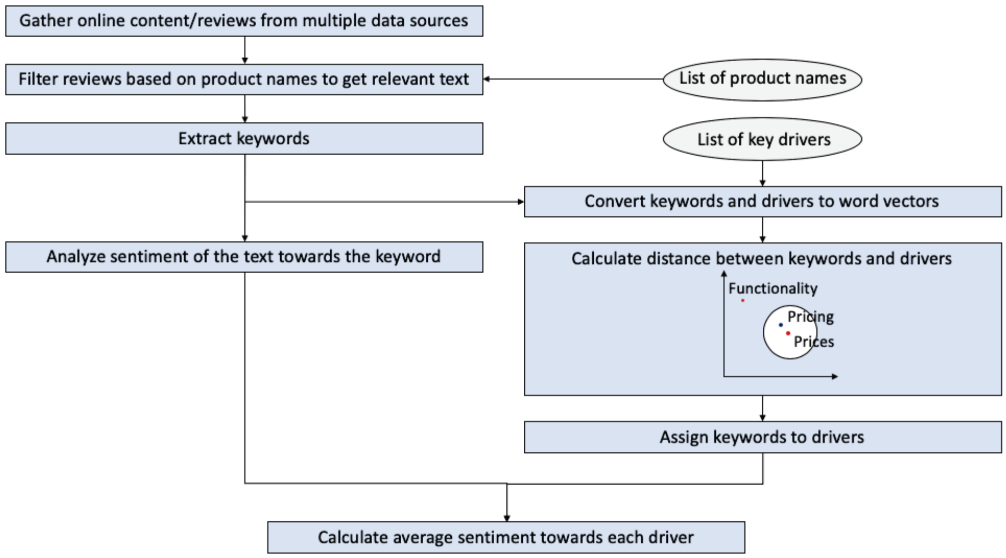

Figure 2: NLP architecture of *Clarity*

convert the keywords as well as the dimensions to word vectors. After the words are converted to word vectors, distance measures (e.g. cosine distance) are used to determine which keywords are most related to which driver. Distance metrics, as well as tuned thresholds, are used to assign keywords to particular dimensions. After assignment, the average sentiment is calculated for each dimension.

The system will then score each product on the factors considered to contribute to performance - sentiment towards key dimensions, star ratings and volume of reviews. The system scores each factor using its percentile score, computed using the average value of the factor for a particular product x compared against the entire distribution of values of that factor for all products. This percentile score will serve as a score for how that particular factor is performing as compared to the competition. An illustration of this percentile scoring mechanism is outlined in Figure 3. In the figure, a distribution of the star ratings is shown for last 18 months for a given data source d_i. Now, given the average star rating (SR) in the time frame of t_n for product x_n is 4.5, we can find the percentile value based on the distribution. Using percentiles makes all the factors of the same scale along with providing an understanding of comparative value of product x_n among all products.

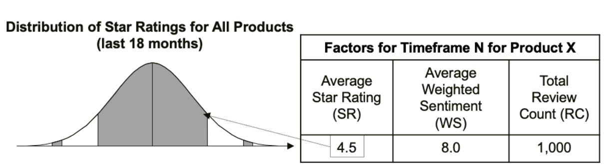

Figure 3: An illustration of the percentile scoring mechanism described in step 8 of Algorithm 2

The system then aggregates across factors to achieve one combined score across all key factors. Percentile scores for each data source are then combined via a weighted sum, where the weights represent the count or volume proportion for that particular data source, as compared to the other data sources. Use of percentiles scores account for the difference in the distributions of the key factors among different data sources. This process will be performed various times using different time frame windows. To reflect the changing product life cycle phases, scores can be computed on 3-month frames, dating back as far back as 18 months. Then to combine across time linear weights can be used, to weight more recent time frames higher than past frames. After aggregating across time frames, the system will output 1 score per product, which holistically represents its competitiveness across the key factors considered. The system can also show the score over time.

Visualization

To better visualize the changes of the scores over time, we leverage *stream graphs* (Byron and Wattenberg 2008). In order to emphasize the changes in ranking over time, we used a variant of the stream graphs called *sorted stream graphs*, a.k.a. area bump charts (see Figure 1), which more appropriately convey the intended insights.

The x-axis, naturally, represents time; each stream represents a product; and the height of each stream represents the score of the product at that particular point in time. The visualization is highly interactive, allowing the user to highlight a stream to better understand the historical changes of its ranking. Also, hovering over the streams brings up a *tool tip* displaying the exact score at that point in time as well as the deltas for the percentage change of the score and change in ranking with regards to previous month.

This visualization is presented in the context of a web dashboard. The dashboard also showcases other visuals that

Algorithm 2: *Clarity* Main Steps

Input: Online content/reviews
Output: *Clarity Score* - a numerical value to denote perceived competitiveness of a product
1. Gather reviews from source d_1 to d_N
2. **for** *analyzing product* x_i **do**
    3. **for** *each 3 month time frame* t_i*: i from 1 to 6* **do**
        4. **for** *each data source* d_i*: i from 1 to N* **do**
            5. Pre-process the text to retrieve average star rating (SR)
            6. Calculate the review count for each product (RC)
            7. Calculate sentiment towards each driver by passing the text through NLP engine described in Figure 2 and combine to calculate weighted average sentiment (WS)
            8. Compute the percentile value of each of the factors (SR, RC and WC) for a particular product x_i based on the distribution of values of that factor for all products x_1 to x_n
            9. Combine percentile scores for SR, RC and WC for a product
        end
        10. Aggregate across all the data sources using a weighted sum where the weights represent the count or volume proportion for that particular data source
    end
    11. Combine across all time frames (linear decay used), to weight more recent time frames higher than past frames
end
12. **for** *each month* **do**
    13. Update the Clarity score for each existing product
end

convey detailed information about the topics and themes extracted and their frequency. All the dashboards visualizations, including the one described in this article, were created using Data-driven documents (D3) (Bostock, Ogievetsky, and Heer 2011)

The method uses commercial data source and in-house data sources and can easily scale to others available over time.

Programmatic interaction/APIs

Along with the visualization, *Clarity* supports invocation via Application Programming Interfaces (APIs). The system provides a collection of REST endpoints to achieve full functionality in a programmatic way i) pull *Clarity Scores* for different products at both individual factor level and aggregated level ii) filter by time and iii) suggested competitors set. Security is a concern and authentication via tokens is required to interact with the API. This enables the core capability of data-driven comparison of products to be integrated and reused in various applications. A few where this has been already completed are: detection of product related events in news media, product pricing recommendation, and talent management. This integration of *Clarity* into other business applications programmatically is accessed by additional hundreds of users daily.

System Improvement and Maintenance

To provide the best text analytics performance, the NLP engine in *Clarity* is always under constant improvement. The original NLP engine in *Clarity* clustered user reviews using a predefined taxonomy. The next step of the improvement is to provide training data to develop a supervised machine learning model. Due to the ambiguity of language understanding, carefully curated high quality data is collected from domain experts. Using the initial data, a machine learning model is developed to provide predictions. Subsequently, the predictions are shown to users, such that new labels can be collected if the users provide feedback on the correct output. With the self sustaining data collection pipeline, the underlying machine learning model is updated.

The main advantage of using a supervised learning model is to classify user reviewers based on the drivers that are of interest to the users. The current implementation using the 5 drivers gives a high level overview of the competitiveness of the products. However, eventually, the model needs to provide more detailed information about specific areas so that the users can devise concrete steps to improve their product's competitiveness. Thus, in the improved model, we expand the 5 drivers to 14 topics, including documentation, performance, etc.

With all the on-going changes in *Clarity*, the program must be maintained properly. With the complexity of the project, lead engineers and scientists take ownership of each part of the program, and to ensure the uninterrupted roll outs of new features and functionalities. E.g., when a new algorithm for the *Clarity Score* is updated, we ensure that the historical values transition properly to the new values.

Apart from the upgrades to the model, the data sources are refreshed independently at different frequencies. To maintain and validate the system's quality, we update the *Clarity Scores* monthly. The model could be refreshed anytime the data is changed. Again, to control system behavior, we update it in controlled released cycles.

Evaluation and Usage Experience

We now discuss the evaluation of *Clarity* in the lab as well as its experience in the field. To test the output of the system, we consider (a) the accuracy of its classification of online reviews and (b) its ranking of products against an alternative ranking commercially available. To test the system's experience, we report on its Net Promoter Score (NPS) and adoption experience.

In-lab Evaluation

Quantitative Evaluation *Clarity* aims to automate the comparative analysis by performing text analytics and machine learning on online reviews. To provide useful business insights, accurate models are necessary. However, because of the ambiguous nature of natural languages where an opinion is usually not necessarily agreed upon by many people, carefully chosen quantitative metrics are employed to measure the accuracy and performance of the implemented models.

One of the main tasks of the *Clarity* model is to classify reviews into topics relevant to business users. Originally, the reviews are categorized into the 5 main drivers. However, this set of categories could be expanded to provide more detailed information. One of the major improvement is to expand the 5 drivers into 36 subtopics and 7 main topic groups, where subtopics grouped into topics. Given the larger set of categories, unsupervised learning approaches are becoming less effective, since the online reviews need to be clustered and categorized in a specific way. Therefore, a supervised learning approach is employed.

To provide training data and establish a more rigorous quantitative evaluation, we implemented a tagging website where a broad audience labels the reviews based on the above pre-defined set of topics and subtopics. Using the collected data, we trained supervised learning models, where the overall classification performance is 0.52 in the micro-averaged f1 score. To understand the performance, we developed a set of ground truth data for testing. Using this new set of data, we requested manual labeling from the same set of audience, and calculated the human classification performance of 0.55 in the micro-averaged f1 score. The reason for such a low score is due to the fact that topic classification is a non-trivial process where different people can label the same sentence very different, depending on their understanding. On the same set of data, our model has a micro-averaged f1 score of 0.51, which is close to the score in the previous dataset, and a score that is near human performance. As noted earlier in the system improvement and maintenance section, our underlying model is always improving to provide the best performance. It is believed that in the next iterations of the underlying NLP model, human performance will be out-performed.

Evaluation by Experts

Gartner Magic Quadrants The IT consulting firm *Gartner* periodically produces a series of market research reports where they rate vendors according upon two criteria: *completeness of vision* and *ability to execute* (Gartner, Inc. n d). Each of these reports include a 2x2 matrix chart similar to the one depicted in Figure 4.

Vendors with both a high completeness of vision and a high ability to execute are called *leaders* (vendors A and E in Figure 4) whereas vendors with low scores in both dimensions are called *niche players* (vendor B). Vendors with a high completeness of vision but with a low score in the ability to execute are called *visionaries* (vendors C and G) and the vendors with a poor completeness of vision but good ability to execute are called *challengers* (vendors D and F).

In the context of 5 different markets randomly selected $\mathcal{M} = \{M_1, M_2, ..., M_5\}$, we examined the results provided by Gartner and noted the similarities and differences with the *Clarity Scores*. Given a Gartner Magic Quadrant

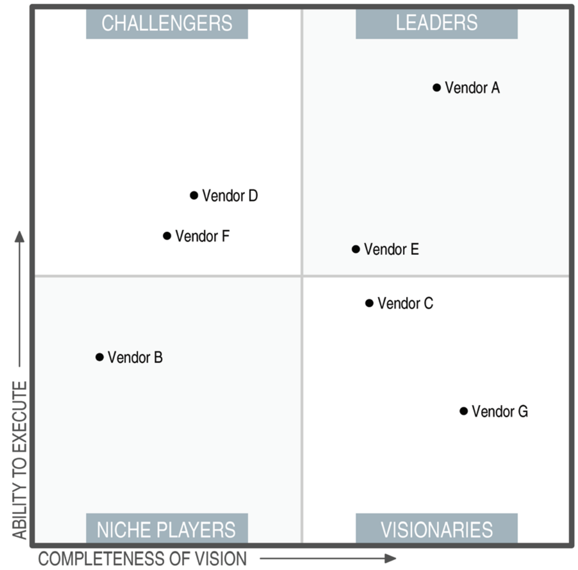

Figure 4: Gartner Magic Quadrant

report for market M_i, we identified the vendors $\mathcal{V}^{M_i} = \{V_1^{M_i}, V_2^{M_i}, ..., V_n^{M_i}\}$ in that report and we then identified the products $\mathcal{P}^{V_j M_i} = \{P_1^{V_j M_i}, P_2^{V_j M_i}, ..., P_m^{V_j M_i}\}$ such as $P_k^{V_j M_i}$ is a product in the market M_i provided by the vendor $V_j^{M_i}$. We then compared the *Clarity scores* for the products in $\mathcal{P}_{V_j}^{M_i}$ for which we had data.

We provide the results of the aforementioned comparison in Table 1. Please note that Gartner provides ratings at a vendor level whereas *Clarity* does it at a product level. For instance, when we analyzed M_3, we considered the two products that the vendor V_1^{M3} provides in that market: $P_1^{V_1 M_3}$ and $P_2^{V_1 M_3}$.

In the set of 5 markets analyzed and with the exception of $P_1^{V_1 M_5}$ (which Gartner ranks as *visionary* but its *Clarity Score* is greater than *Clarity scores* for the products ranked as *leaders* , i.e. $P_1^{V_2 M_5}$ and $P_1^{V_3 M_5}$), (1) and (2) hold true within a given market M_i:

$$\mathcal{CS}(p_L) > \mathcal{CS}(p_C) > \mathcal{CS}(p_N) \tag{1}$$

$$\mathcal{CS}(p_L) > \mathcal{CS}(p_V) > \mathcal{CS}(p_N) \tag{2}$$

where $\mathcal{CS}(p_i)$ is the *Clarity Score* for product p_i and p_L, p_C, p_V, p_N are products whose vendors are ranked as leaders, challengers, visionaries, and niche players respectively. This evaluation suggests that *Clarity Scores* and Gartner scores are aligned when it comes to assessing competitiveness.

Net Promoter Score® (NPS)[2] The *Net Promoter Score* is a metric used in customer satisfaction research. The NPS is

[2] Net Promoter, Net Promoter System, Net Promoter Score, NPS and the NPS-related emoticons are registered trademarks of Bain & Company, Inc., Fred Reichheld and Satmetrix Systems, Inc.

Market	Vendor	Product	Gartner Q	*Clarity* score
M_1 = Access Management	$V_1^{M_1}$	$P_1^{V_1 M_1}$	Leader	8.01
	$V_2^{M_1}$	$P_1^{V_2 M_1}$	Leader	7.68
	$V_3^{M_1}$	$P_1^{V_3 M_1}$	Leader	4.04
	$V_4^{M_1}$	$P_1^{V_4 M_1}$	Visionary	3.9
	$V_5^{M_1}$	$P_1^{V_5 M_1}$	Visionary	3.12
M_2 = Ind. IoT Platf.	$V_1^{M_2}$	$P_1^{V_1 M_2}$	Visionary	3.96
	$V_2^{M_2}$	$P_1^{V_2 M_2}$	Niche player	3.41
M_3 = Cloud IaaS	$V_1^{M_3}$	$P_1^{V_1 M_3}$	Leader	8.93
	$V_1^{M_3}$	$P_2^{V_1 M_3}$	Leader	8.72
	$V_2^{M_3}$	$P_1^{V_2 M_3}$	Leader	8.43
	$V_3^{M_3}$	$P_1^{V_3 M_3}$	Leader	8.42
	$V_4^{M_3}$	$P_1^{V_4 M_3}$	Niche Player	7.86
M_4 = Enterprise Agile Planning	$V_1^{M_4}$	$P_1^{V_1 M_4}$	Challenger	8.83
	$V_2^{M_4}$	$P_1^{V_2 M_4}$	Challenger	8.21
	$V_3^{M_4}$	$P_1^{V_3 M_4}$	Niche Player	6.18
	$V_4^{M_4}$	$P_1^{V_4 M_4}$	Niche Player	5.96
	$V_5^{M_4}$	$P_1^{V_5 M_4}$	Niche Player	3.94
M_5 = Cloud Financial Planning and Analysis	$V_1^{M_5}$	$P_1^{V_1 M_5}$	Visionary	8.77
	$V_2^{M_5}$	$P_1^{V_2 M_5}$	Leader	8.22
	$V_3^{M_5}$	$P_1^{V_3 M_5}$	Leader	7.08
	$V_4^{M_5}$	$P_1^{V_4 M_5}$	Visionary	4.79
	$V_5^{M_5}$	$P_1^{V_5 M_5}$	Visionary	4.04

Table 1: Comparison of ratings provided by Gartner vs. *Clarity Score*

calculated based on responses to the question: *How likely is it that you would recommend this product?* and the answer is based on a 0 to 10 scale. People who have responded with a score of 9 or 10 are called *promoters*, those who have responded with scores between 7 and 8 are called *passives*, and the ones that have responded with scores between 0 and 6 are called *detractors*. The NPS is the difference between the percentage of promoters and detractors. Therefore, NPS range is -100 to 100. In order to validate *Clarity Score* we analyzed NPS data for 476 products. We only considered the 50 products that had at least 150 responses to ensure the data is not biased. In Figure 5, we created a scatter plot for the 50 products: the horizontal axis shows *Clarity Scores* and the vertical axis represents NPS.

From the plot, it's clear that, with the exception of a few outliers, high values of NPS correspond to high values of *Clarity Scores*, showing that *Clarity* is also aligned with NPS for competitiveness assessment.

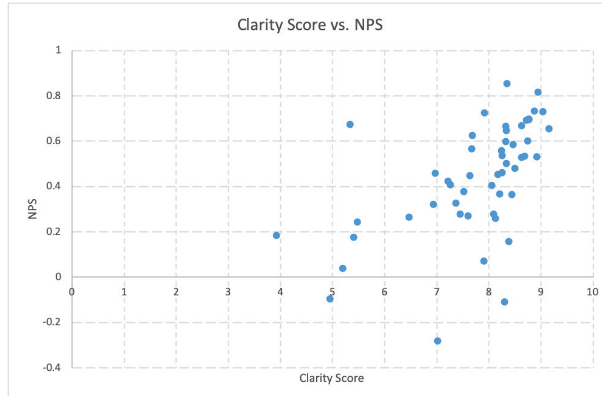

Figure 5: Comparison of *Clarity Score* vs. NPS for 50 products

In-field Evaluation

We have gathered user feedback for in-field evaluation of the system through multiple design thinking workshops and over 50+ user interviews. We specifically wanted to understand how the different divisions in IBM were using the system so we interviewed product managers, market researchers, and executives.

Overall, the users agreed that *Clarity's* competitive analysis aligns with their perception of the products and competitors, with some natural resistance when the ranking was not favorable for their particular product. Users also stated that although they find *Clarity* very useful and it covers the major components of competitive analysis in near real-time, it could be further expanded to showcase additional aspects of competitiveness.

We used *Net Promoter Score*, a metric used in customer satisfaction research, to capture user feedback. The details of the process to calculate NPS is mentioned in the section above. *Clarity* received the NPS score of 52. As mentioned, the range of NPS is -100 to 100 and therefore 52 represents a strong positive opinion among the users for *Clarity*.

Based on the feedback of the users we are continuing to enhance our back-end NLP engine. Some of the system improvements are discussed under system overview.

Adoption Experience

The system has been deployed for over a year and currently provides over 160 competitive analyses involving data from over 800 products. It is being used by over 1,500 users globally in market intelligence, product development and sales, with a steady use over 16K unique views per quarter.

Business Impact

Clarity provides a near real-time competitive insights and recommendation to a wide range of users. This is a significant use-case of leveraging NLP and process automation to change an existing manual, laborious, sometimes biased and not so repeatable practice to an automated, intelligent, and sustainable system. *Clarity* has been augmenting end-users subject matter expertise and knowledge to improve their decision making process leveraging AI.

Discussion - *Clarity as a Deployed AI System*

In this section, we discuss the characteristics of *Clarity* as required for a Deployment track paper but not discussed elsewhere. *Clarity* is a data-driven, unsupervised system to compare products in a market place using user-generated opinions from public data sources. Its can be invoked for human consumption to get a rich visualization and also for programmatic integration into other applications via APIs. It has been extensively evaluated for the quality of its output and usability, and found to be state-of-the-art as well as cost-effective.

Development and Deployment

The development of *Clarity* started in Oct 2017 with 8 developers and the first system was deployed in a year (Oct 2018). The system has been continuously evaluated and its components updated based on business needs.

One main challenge during development was access to clean data related to the products. Although automated approaches were explored in the beginning, the cleanest data was achieved through different collaborations and maintaining relationships with data providers. Another challenge was lack of labeled data, which forced us to explore unsupervised techniques. This was further challenging due to the ambiguous nature of natural language especially in the textual content written in reviews. These challenges were addressed by tuning the thresholds of the models. After deployment, new challenges were raised. One challenge was evaluating the accuracy of the system especially since opinions of performance differed among experts. We addressed this by using carefully chosen quantitative metrics which showcase the accuracy and performance of the implemented models. Another challenge was comparing our perceived competitiveness score of a product with other external market experts.

Usage of AI technology

Clarity uses AI technology extensively in its processing of user-generated content. (1) It performs information extraction, sentiment analysis and semantic analysis using the commercially available IBM Watson Natural Language Understanding (NLU) service. (2) The system also relies on word vectors for NLP tasks. A variety of word vectors have been used: pre-trained Word2Vec, custom trained models using online reviews of supported tools and a mix of word vectors. (3) Clustering techniques are used to verify *Clarity* analysis drivers and new ones are considered for inclusion.

The system also uses AI methods for generating visualization. (1) Rules to track explanations and ordering of products in results, and (2) Formatting of results based on cognitive considerations.

Conclusion

In this paper, we considered the problem of comparing products in a marketplace automatically from online content. This is an important business activity that marketers, sellers and product managers conduct regularly. Unfortunately, it is also very time consuming and costly which can be particularly challenging for businesses with large product portfolios and fast-changing customer environment.

In response, we presented *Clarity*, an unsupervised data-driven system for assessment of products and its deployment in a large IT company. The system has been running for over a year and used by over 1500 people performing over 160 competitive analyses involving over 800 products. The system performs NLP methods on a collection of online content and computes competitive results. The results (scores) and explanations of factors leading to the results are visualized in an interactive dashboard that allows users to track their product's performance as well as understand main contributing factors. *Clarity* has thus proven to be an excellent example of an AI-based system that has been integrated and reused in various applications such as product pricing recommendation, and talent management and has performed extremely well in critical business activities.

References

Afful-Dadzie, E.; Nabareseh, S.; Oplatková, Z. K.; and Klimek, P. 2014. Enterprise competitive analysis and consumer sentiments on social media - insights from telecommunication companies. In *DATA*.

Amarouche, K.; Benbrahim, H.; and Kassou, I. 2015. Product opinion mining for competitive intelligence. *Procedia Computer Science* 73:358 – 365. International Conference on Advanced Wireless Information and Communication Technologies (AWICT 2015).

Bhatt, D. A.; Mcneil, K. E.; and Patel, N. A. 2014. Time-based sentiment analysis for product and service features. In *Journal Software, VOL. 9, NO. 2,*.

Bostock, M.; Ogievetsky, V.; and Heer, J. 2011. D3: Data-driven documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*.

Byron, L., and Wattenberg, M. 2008. Stacked graphs – geometry & aesthetics. *IEEE Transactions on Visualization and Computer Graphics* 14(6):1245–1252.

Devlin, J.; Chang, M.; Lee, K.; and Toutanova, K. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR* abs/1810.04805.

Gartner, Inc. n. d. Gartner magic quadrant research methodology. https://www.gartner.com/en/research/methodologies/magic-quadrants-research. Accessed: 2019-09-05.

Joung, J.; Jung, K.; Ko, S.; and Kim, K. 2018. Customer Complaints Analysis Using Text Mining and Outcome-Driven Innovation Method for Market-Oriented Product Development. *Sustainability* 11(1):1–14.

Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; and Sutskever, I. 2018. Language models are unsupervised multitask learners.

Srivastava, B.; Chetlur, M.; Gupta, S.; Vasa, M.; and Visweswariah, K. 2018. Effective business development for in-market it innovations with industry-driven api composition. In *Data Science Landscape. Studies in Big Data, Vol 38. Springer*.