

RESEARCH

Open Access



DERIV: distributed brand perception tracking framework

Manu Shukla^{1*} , Raimundo Dos Santos², Andrew Fong¹ and Chang-Tien Lu³

*Correspondence:

mashukla@vt.edu

¹ Omniscience Corp, Palo Alto, CA, USA

Full list of author information is available at the end of the article

Abstract

Determining user's perception of a brand in short periods of time has become crucial for business. Distilling brand perception directly from people's comments in social media has promise. Current techniques for determining brand perception, such as surveys of handpicked users by mail, in person, phone or online, are time consuming and increasingly inadequate. The DERIV system distills storylines from open data representing direct consumer voice into a brand perception. The framework summarizes perception of a brand in comparison to peer brands with in-memory distributed algorithms utilizing supervised machine learning techniques. Experiments performed with open data and models built with storylines of known peer brands show the technique as highly scalable and accurate in capturing brand perception from vast amounts of social data compared to sentiment analysis.

Keywords: Big data, Distributed learning, Brand perception, In-memory distribution

Introduction

There is a glaring need to improve the tracking of brand perception which is ill-served due to time-consuming techniques [1]. Online surveys take pre-set questions from companies and present them to users. Offline surveys handpick representative users and ask them detailed questions about products. Responses are then carefully analyzed making the process time consuming and expensive. The cumbersome nature of traditional survey techniques also preclude companies from taking advantage of new trends or rapidly rectifying negative developments in perception. This work presents DERIV, a novel framework to track user perception of a brand in near real time using open data such as tweets. Current techniques that measure brand perception rely on the sentiment of users. This approach is limited as most opinions from customers have little or no sentiment attached to them. For instance, the phrase 'Electric Car Z goes 300 miles in a single charge' shows positive sentiment towards a brand. However, sentiment analysis techniques will frequently identify this statement as neutral. Measuring sentiment from each customer tweet or social media post also does not convey what is being said about a brand across sources and across elements. Instead of using raw social media posts, we employ storylines (see the next paragraph for an example of what a storyline looks like) which are entities (people, organizations, things) linked by edges represented by the observed relationships between the entities. These relationships are normally the verbs

extracted from text preceding the following entity. Storylines not only combine consumer voice across tweets, review sites, and consumer forums, but also connect them over multiple entries. The key reasons for measuring user perception from storylines are:

1. *Connect user voice across data elements and sources* Each data element such as tweets by themselves only provide isolated cases of users interaction with a brand. Storylines, on the other hand offer a comprehensive view of users' perception of a brand.
2. *Analyze entities with multiple relationships* The storylines are a compact way to represent the interactions among entities through multiple relationships.
3. *Eliminate noise from consumer voice* Connecting entities into storylines eliminates the clutter of slang words and brand-specific terms, and unclear verbiage that is prevalent in tweets.

As an example of how storylines are generated consider the following. User "A" tweets during an election that "Candidate X is the new #koch brothers darling!" and user "B" tweets that "Unfortunately, #Koch brothers only #support the #Establishment who will do their bidding like #Billionaires supporting #Hillary!". A possible storyline, then, would be 'candidate X → new #koch brothers darling! → #support the #establishment → #billionaires supporting #hillary'. This storyline connects the entities across the two tweets, and their combination better represents impact of a negative subject (the establishment) at the time on the brand (candidate X) than each individual tweet does.

Finding perception of a brand from social media postings creates several issues:

- | | |
|-------------|---|
| Challenge 1 | Scaling to large number of postings. In social media, the number of mentions of a brand or topic can be extremely high on a daily basis, anywhere from hundreds to millions. Processing this amount of data from storylines and calculating brand perception requires ability to scale. |
| Challenge 2 | Summarizing the storylines into a trackable measure. Since the comments on social media can be brief and fractured, condensing disparate elements into a meaningful trackable score is difficult. |
| Challenge 3 | Calculating perception uniformly across brands. Generating a perception measure that is consistent across brands over time is critical. |
| Challenge 4 | Validating the measure on several types of brands. Since brands can vary widely, ensuring that the perception measure works consistently across large spectrums is key to trustworthiness of the measure. |

This paper proposes a machine learning based framework to consistently measure the perception of a brand from social media in comparison to its peer brands. It utilizes in-memory distribution to build and update models and scale to amounts of data on social media to generate a consistent perception in short time periods. The key contributions of the paper are:

1. *Novel model to calculate brand perception from storylines* Our model combines multiple classifier scores from large number of storylines to distill a comprehensive perception of a brand. This perception is calculated from the true voice of customer.

2. *Innovative categorical bands to consistently measure perception against peer brands*
Allocating brands storylines into perception bands allows consistency in measure across peer brands.
3. *Distributed algorithms to perform supervised learning and scoring at scale*
The algorithms for perception modeling use distributed in-memory techniques that scale the building of multiple classifiers with increasing number of labeled storylines and scoring large number of new storylines about a brand for a time period.
4. *Extensive experiments to validate the perception scores*
Using Twitter data on several brands from diverse categories to perform experiments, the results show the relevance and effectiveness of our calculated perception as compared to sentiment analysis.

The rest of the paper is organized as follows. “[Related works](#)” section elaborates on the existing work in the areas of brand perception based on sentiment analysis and machine learning. “[DERIV brand perception](#)” section provides details of the approach in DERIV and the perception modeling. In “[DERIV system](#)” section, an overview of DERIV architecture and key-value pair based distribution algorithms for perception modeling is provided. Experiments and use cases are described in “[Experiments](#)” section and the overall conclusions presented in “[Conclusions](#)” section.

Related works

In the realm of brand perception, the scientific literature can be divided into three distinct domains.

Marketing based brand perception measurements

Marketing-oriented perception measures have been researched for a long time. Social perception theory has been used to measure brand perception [2]. Cultural dimension and social influence on brand perception has been examined [3]. Impact of brand perception on luxury item purchases has been explored [4]. Connection between quality and perception of a brand has been investigated [5]. Users sense of a brand has also been studied [6]. Users selection of a brand based on multiple factors has been explored [7]. Experiments with high-share brands loyalty have been described [8]. Effect of shape on brand perception has been discussed [9]. Impact of celebrity on brand perception has been investigated [10]. Semantic analysis of web data to determine reputation of a brand against its competitors has been performed [11]. This corpus of work underscores the need to calculate brand perception and study impact of various factors on perception. However these measures are usually relative and can not be performed over short periods.

Sentiment analysis

Sentiment analysis has been used extensively to measure personal feelings towards a brand. Opinion mining has been surveyed [12]. Extracting sentiments from tweets has been explored [13] while brand sentiment analysis has also been studied [14]. Text mining techniques have been used to determine user sentiment towards well known brands [15]. Detecting polarity in tweets helps in gauging customer sentiment towards a brand

[16]. Classifier ensembles have been explored for tweet sentiment analysis [17]. Machine learning techniques have been applied to perform sentiment based opinion mining [18]. The sentiment analysis techniques however focus on individual tweets and have no ability to detect perceptions by connecting user voice across sets of multiple tweets. They are also easily fooled by sarcasm. Hence these techniques are better suited for user comments monitoring rather than tracking perception.

Social media mining

Social media mining has been a rich source of information on brands [19]. Use of social media for knowledge acquisition and validation is well known [20]. Linking new articles to generate evolving new stories is popular [21]. Interactions of storylines in news has been explored [22]. Building storylines of text, pictorial and structured data has been investigated [23]. Storylines have been used to determine evolving events effectively [24]. Social media mining has not been used to measure brand perception so far.

None of these techniques can measure the perception of a brand compared to its peer brands on a daily or shorter period basis. In addition, storylines from open data to the best of our knowledge have never been used to measure brand perception. Unlike DERIV, other measures do not combine user voice from multiple data elements and multiple sources into a single trackable entity.

DERIV brand perception

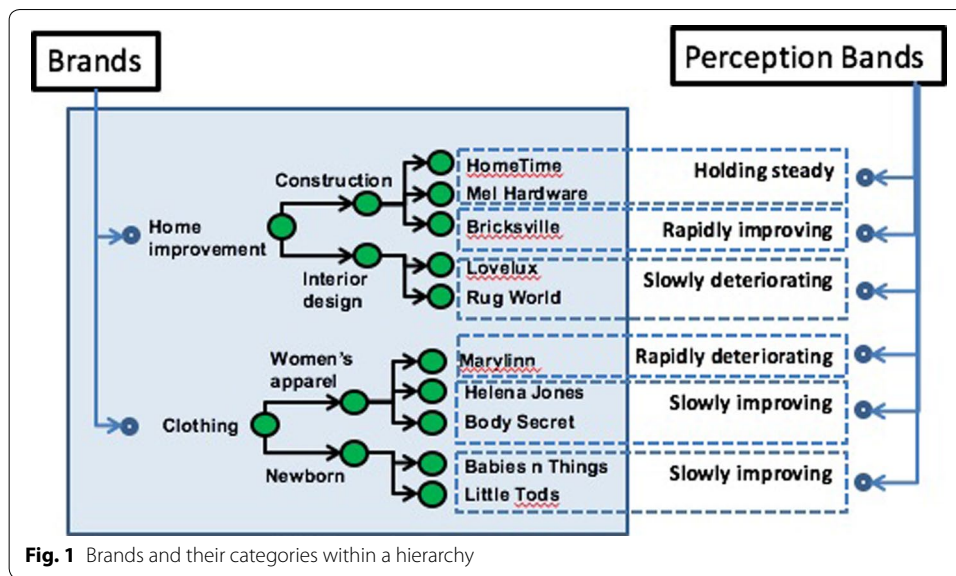
This section describes the modeling and scoring techniques used to generate the DERIV perception. “[Perception bands](#)” section provides overview of the bands in which perception is slotted into and “[DERIV flow](#)” section provides overview of steps in perception calculation, bands, a brand and its hierarchy, modeling and scoring used in brand perception. “[Comprehensive perception modeling](#)” section details the measure calculation and “[Root cause analysis](#)” section details root cause analysis done to identify key postings that tie to perception.

Perception bands

The bands are pre-determined slices of perception in which the storylines related to a brand are placed based on pre-existing survey data or industry standard measures (see Fig. 1 for examples of bands).

Sentiment analysis has the limitation that sentiment of a post or a tweet can only be categorized as positive, negative or neutral. This, however, does not show the trajectory of a brand, which is more relevant to measure, rather than observing its state at a given time [25]. Only then can we understand the movements in brand perception and the factors influencing those movements. That motivated us to create five bands to capture trajectories of a brand’s perception. The bands used in this study are:

1. Rapidly improving (RI)—This band puts brand at a rapidly improving path of perception irrespective of its current perception level.
2. Slowly improving (SI)—Brands in this band have improving perception from any level.

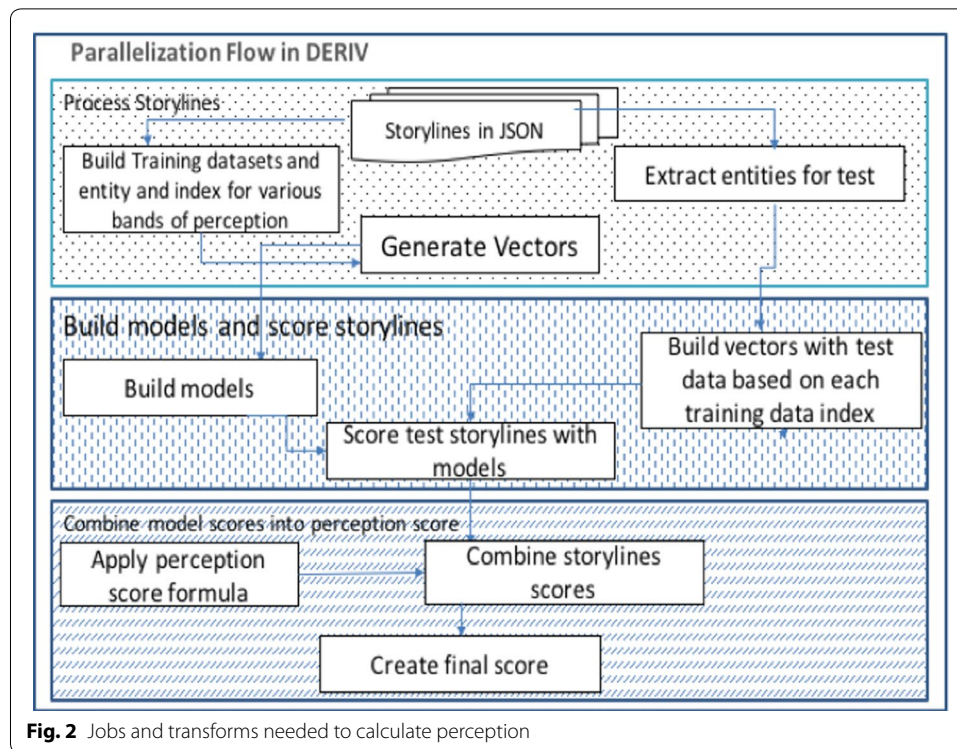


3. Holding steady (S)—Models stable brand perception. That suggests that the perception of a brand is not moving much in either positive or negative direction but instead maintaining its state.
4. Slowly deteriorating (SD)—This means the perception of a brand has started to head downwards from its previous state.
5. Rapidly deteriorating (RD)—This band indicates the perception of the brand has started to fall rapidly in the eyes of consumers.

When a storyline has a favorable take on a brand, it is added to the set that represents either “slowly improving” or “rapidly improving” bands. Similarly, unfavorable storylines and storylines showing brand stability in perception will be added to the “rapidly deteriorating”, “slowly deteriorating” or “holding steady” perception bands models. Storylines added to a band are labeled positive if they fit the favorable/unfavorable/steady perception. All of the above are considered positively-labeled training data set. Negatively-labeled elements are the ones that do not contribute to the respective brands one way or another (cannot determine if favorable or unfavorable).

DERIV flow

DERIV employs storylines and models built with supervised learning techniques to generate perception. Figure 2 shows the flow of data and operations in DERIV. The first block in the figure shows storylines being processed in parallel to generate vectors that in the subsequent block are used to build models. *N* models are built, one for each of the *N* bands followed by scoring the test storylines in parallel against each of the models. The storylines are a sequence of entities and relationships and are treated as bag of words documents for both training classifiers and scoring against them. Finally the positive scores above a threshold δ in each band are counted and their counts are used to calculate comprehensive perception in final block.



The brands are associated with a category which can be organized in a hierarchy. The category “big box stores” can have sub-categories such as “multi-brand retailers”, “construction and home”, etc. An example of brands and their hierarchies is shown in Fig. 1. Each category has the same N bands and a peer brand associated with the band. The peer brands for which storylines are generated are picked from the same or adjoining sub-categories. Storylines for each of the bands can be picked from any peer brand’s storylines. The specific brand names are removed from storylines before training a model with them. Relevant storylines for each brand are assigned to a band’s training data by an analyst. The analyst also determines filtering keywords to collect data for a brand and its peer brands. The collected data is subsequently used in the generation of storylines from which the analyst labels them as representative or non-representative with respect to that brand’s perception within the band. Some examples of perception bands and brands in them are shown in Fig. 1. Training data is used to build classifiers that capture binary-class labeling for each brands storylines as representative of being in the perception band or not. N of these classifiers, one for each band, are combined into a model. The combination of Support Vector Machines (SVM) classifiers is used to build the final model to generate comprehensive perception. Every storyline for the brand whose perception needs to be calculated is scored against each of the classifiers. The band with the highest positively scored storylines above threshold determines the band perception will be placed in and the counts for other bands are used to tweak perception further within the band.

Comprehensive perception modeling

The perception is based on the number of positively-labeled storylines for a given time period and represent one of the N bands. The classifiers for each of the bands in

“Perception bands” section give a score to the storyline. The final perception is then calculated using those scores according to one of the five cases described further below. Note that for each case, its corresponding formula ensures that the final score has the following ranges: [80–100] when “rapidly improving” storylines are maximum among the ones that meet the classifier threshold; [60 and <80] when “slowly improving” storylines are the maximum; [40 and <60] when “stable” band storylines are of maximum count, [20 and <40] for “slowly deteriorating” and [0 and <20] for “rapidly deteriorating” storylines with maximum count. These ranges are for the 5 bands as described in “Perception bands” section and can be adjusted so that it makes sense on a per-application basis. We use the following five cases:

1. Highest number of storylines that score above δ threshold are labeled as “rapidly improving”:

$$S_{RI} == \max S_i; C_S = \frac{S_{RI}}{\sum S_i} * A_b + Base_{RI} + A_c * \frac{S_{SI}}{\sum S_i} + A_d * \frac{S_S}{\sum S_i} - A_d * \frac{S_{SD}}{\sum S_i} - A_d * \frac{S_{RD}}{\sum S_i} \quad (1)$$

Equation 1 models perception score such that if maximum number of scored storylines lie in the rapidly improving band, then the overall score stays in the rapidly improving range. The overall score is still penalized by the number of storylines scoring in the two deteriorating perception bands. In addition, the overall score is supplemented by slowly improving and stable band storylines count. This formula attempts to incorporate the concept that trajectory of a brand’s perception is impacted more by adjoining trajectories and less by inverse trajectories as there is a natural progression in the users perception towards a brand. In order to supplement and penalize such that the overall score remains within 0 and 100, the coefficients A_b , A_c and A_d are applied for the particular band and adjacent bands. A_b is 20 when N is 5, A_c is 10 and A_d is 5. The coefficients A_c and A_d become negative when applied to deteriorating bands. When choosing the values of coefficients, it was assumed that there was separation between storylines scoring above the threshold by band. Thus the ones that score above the threshold for slowly improving do not score above threshold for any other band. The range of final score within band ranges is mathematically enforced only if there is good separation in the storylines for the positive labels for each band. When maximum number of storylines is rapidly improving, the number of storylines above threshold is anywhere from a fifth to all the storylines for rapidly improving perception band. That makes the contribution of those storylines to the final score between 4 and 20. When number of rapidly improving storylines is high compared to other bands, their increment to $Base_{RI}$ is close to 20, while that of other bands is close to 0 which brings the final score close to 100. When there are high numbers of deteriorating band storylines above threshold, then their maximum decrement to the final score is 2, while the minimum increment is 4 for the rapidly improving band, and 2 and 1 for slowly improving and stable bands, bringing the score closer to the base rapidly improving mark.

2. Highest number of storylines that score above δ threshold are labeled as “slowly improving”:

$$S_{SI} == \max S_i; C_S = \frac{S_{SI}}{\sum S_i} * A_b + Base_{SI} + A_c * \frac{S_{RI}}{\sum S_i} + A_c * \frac{S_S}{\sum S_i} - A_d * \frac{S_{SD}}{\sum S_i} - A_d * \frac{S_{RD}}{\sum S_i} \quad (2)$$

The final measure is calculated using the formula in Eq. 2. Similar to Eq. 1, the score for slowly improving band is supplemented by rapidly improving and stable bands storylines, but penalized by slowly deteriorating and rapidly deteriorating band storylines.

3. Highest number of storylines that score above δ threshold are labeled as “stable”:

$$S_S == \max S_i; C_S = \frac{S_S}{\sum S_i} * A_b + Base_S + A_d * \frac{S_{RI}}{\sum S_i} + A_c * \frac{S_{SI}}{\sum S_i} - A_c * \frac{S_{SD}}{\sum S_i} - A_d * \frac{S_{RD}}{\sum S_i} \quad (3)$$

Here, slowly improving and slowly deteriorating bands are considered adjacent bands and their storyline counts weigh higher while storyline counts of rapidly deteriorating and rapidly improving bands are weighed lower.

4. Highest number of storylines that score above δ threshold are labeled as “slowly deteriorating”:

$$S_{SD} == \max S_i; C_S = \frac{S_{SD}}{\sum S_i} * A_b + Base_{SD} + A_d * \frac{S_{SI}}{\sum S_i} + A_d * \frac{S_{RI}}{\sum S_i} + A_c * \frac{S_S}{\sum S_i} - A_c * \frac{S_{RD}}{\sum S_i} \quad (4)$$

In Eq. 4, the score is penalized further by storylines in rapidly deteriorating perception band while raised by storylines count in stable, slowly improving and rapidly improving bands with varying weights. The $Base_{SD}$ weight is added along with the increment for slowly deteriorating storylines in a positive direction to keep the score above 20.

5. Highest number of storylines that score above δ threshold are labeled as “rapidly deteriorating”:

$$S_{RD} == \max S_i; C_S = \frac{S_{RD}}{\sum S_i} * A_b + Base_{RD} + A_d * \frac{S_{RI}}{\sum S_i} + A_d * \frac{S_{SI}}{\sum S_i} + A_d * \frac{S_S}{\sum S_i} - A_c * \frac{S_{SD}}{\sum S_i} \quad (5)$$

The final perception calculated using Eq. 5 is penalized for storylines in slowly deteriorating band, but improved for ones in stable, slowly improving and rapidly improving bands. The increment for rapidly deteriorating storylines is added in a positive direction to keep the score above 0.

In Eqs. 1–5, S_i represents the count of positive scores of the storylines for a given model S . The equations first identify the band within which the score will lie and then fine tune the score based on the count of storylines in the determined brand and in progressive adjacent bands. Based on these principles the final generalized perception equation can be written in a common form for five bands as depicted in Eq. 6.

$$C_S = \frac{S_{maxS}}{\sum S_i} * A_b + Base_{maxS} \pm A_c * \frac{S_{maxS-1}}{\sum S_i} \pm A_c * \frac{S_{maxS+1}}{\sum S_i} \pm A_d * \frac{S_{maxS-2}}{\sum S_i} \pm A_d * \frac{S_{maxS+2}}{\sum S_i} \pm A_d * \frac{S_{maxS-3}}{\sum S_i} \pm A_d * \frac{S_{maxS+3}}{\sum S_i} \pm A_d * \frac{S_{maxS-4}}{\sum S_i} \pm A_d * \frac{S_{maxS+4}}{\sum S_i} \quad (6)$$

C_S represents the cumulative DERIV perception. A_i represents the weights assigned to each band's count of positively labeled storylines. The values assigned to the A_i coefficients depend on their adjacency to the band with the highest count. $Base_{RI}$ is 80, $Base_{SI}$ is 60, $Base_S$ is 40, $Base_{SD}$ is 20 and $Base_{RD}$ is 0. Hence the equations are designed such that score C_S greater than $Base_{RI}$ indicates the measured brand has rapidly improving perception, perception greater than $Base_{SI}$ and less than $Base_{RI}$ is slowly improving, perception between $Base_S$ and $Base_{SD}$ is stable perception, perception between $Base_{RD}$ and $Base_{SD}$ is slowly deteriorating and perception lower than $Base_{SD}$ is rapidly deteriorating perception. The equation has negative sign for S_{RD} and S_{SD} for the coefficients A_i when $i \in \{c,d\}$ and positive otherwise.

The formula generates a normalized score between 0 and 100 with 0 being extremely unfavorable brand perception and 100 being extremely favorable. For example, if a brand's open data elements generate 100 storylines that score above threshold for band classifiers, 25 of which score above the threshold for slowly improving, 45 for rapidly improving, 20 for stable and 10 for slowly deteriorating band classifier, then its perception score for the time period will be $80 + (45/100) * 20 + (25/100) * 10 + (20/100) * 5 - (10/100) * 5 = 80 + 9 + 2.5 + 1 - 0.5 = 92.0$.

The formulae synthesize the counts of positively labeled test storylines for each band. The basic band that the final perception lies in is determined by the band with the maximum count of positively labeled storylines of the N bands. The perception is further adjusted within the band by selectively weighing adjacent positive or negative bands higher. It is penalized for high negative perception band label counts and supplemented with positive band storylines count.

Root cause analysis

To tie the movements in brand perception to the actual customer voice the scores were tied back to the data elements they were created from. The top scoring storylines in each band were mapped back to the individual data elements from which the entities in the storylines were extracted. This provides the user the top data elements that are most influential in driving the perception of the brand in each brand. They can then devise a strategy to address the issues causing drop or stabilization in brand perception and divert resources to the factors that are influencing the improving perception bands.

To determine root cause for a perception consider constituent storyline S_a with entities e_i, e_j, e_k, e_l connected together as

$$S_a : e_i \rightarrow e_j \rightarrow e_k \rightarrow e_l$$

where data elements for each entity are

$$e_i \in \{de_b, de_c, de_d, \dots\}, e_j \in \{de_f, de_g, \dots\}, e_k \in \{de_h, de_m, \dots\} \text{ and } e_l \in \{de_n, de_o, \dots\}$$

and the union of each entities data elements for all the storylines associated with the perception band gives the perception's root cause P_{rc}

$$P_{rc} \rightarrow \{de_b, de_c, de_d, de_f, de_g, de_h, de_m, de_n, de_o, \dots\}$$

as data elements behind the storylines of interest for the brand. These data elements give the root cause for the perception generated from storylines that score highly against a particular perception band model. Analyzing the raw data elements and storylines generated from them provides the analysts insights on the media posts behind the perception being in a particular band.

DERIV system

This section presents the architecture of the DERIV framework and provides the details of distributed in-memory algorithms. “Architecture” section describes the architecture of the system and “Distributed algorithms” section describes the algorithms used. The system is designed to be able to scale to updating models rapidly by building models for all the bands with any additional or updated training data for any of the bands. The updated models are then used to score potentially large amounts of storylines generated for the brand having its perception measured.

Architecture

The architecture of DERIV is shown in Fig. 3. The DERIV framework is a sequence of Spark jobs [26] that run on AWS (Amazon Web Services) EC2 (Elastic Compute Cloud) clusters and continues with storylines generated by DISCRN [27] based on traversing ConceptGraph [28]. It proceeds to build models from training data and storylines scores from testing data for a brand whose perception is being calculated. DERIV uses in-memory distribution techniques based on Apache Spark framework that allow computations to be distributed in-memory over a large number of nodes in a cluster [29]. The SVM classifiers used are from Spark MLLib library [30]. The programming constructs available in Spark are reading of data on disk into Resilient Distributed Datasets (RDDs) in-memory and then applying transforms (map, flatMap, filter, reduceByKey, join) and actions (reduce, collect, count) on the RDDs to generate values that can be returned to the application or stored on distributed disk for analysis. Broadcast operation allow for caching variables on each machine of cluster. RDDs provide fault tolerance in case one or more nodes of the cluster fail. The architecture shows the AWS components used by DERIV including AWS EC2 cluster and S3 distributed file store. The modules are divided into two groups.

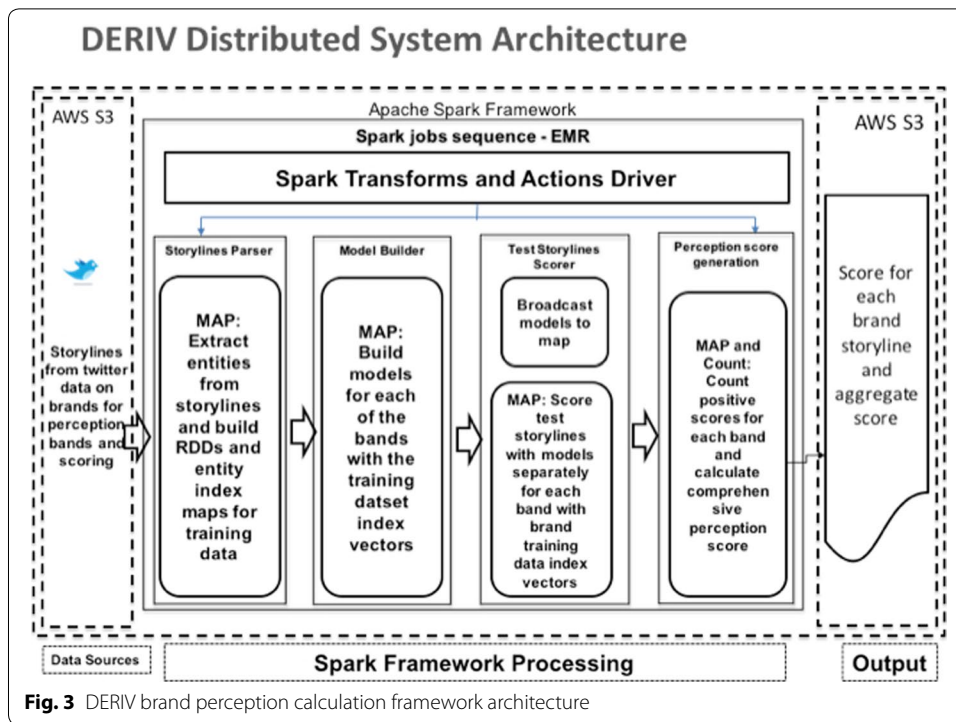


Fig. 3 DERIV brand perception calculation framework architecture

Storyline parsing and classifier building modules

The first module in the architecture flow reads storylines and creates RDDs from training or testing data read from disk and stores a dictionary of storyline terms along with their integer index for the band. The second module creates LabeledData and Vector objects with the indices and builds classifiers with training data RDDs if new training data is provided. The RDD operations for vectorizing storylines and training classifiers are all performed in parallel with storylines modeled as individual observations.

Storyline testing and applying perception models modules

The third module iteratively scores test storylines data against each band’s classifier using the dictionary indexes for training data. It first builds test data vectors with the training data dictionary terms and keeps scores in storylinesResults object’s RDD. The fourth module generates counts of positive storyline scores for each classifier that are above the threshold. It then applies perception model on the counts for calculating the comprehensive DERIV measure.

Distributed algorithms

The algorithm used in DERIV to generate the N classifiers, one for each band for known peer brands and subsequently to generate perception measure with test storylines score from the band classifiers are described in this section.

Build classifiers

Algorithm 1 Generate classifiers from training data

Input: $\{storyline_i\}, \{bands_k\}, svmModels$ {bands and labeled storylines for supervised learning}
Output: $\{bandmodel_k, dictionary_k$ under each $band_k$
 {labeled storylines for a brand under each band}
 1: {step 0: build vectors and index for each band labeled data}
 2: for all band \in bands do
 3: RDD<String> $bandLabeledStorylinesRDD_{band}$ using textfile method to read labeled training data for band on disk
 4: PairRDD<String, Long> $entityIndexRDD_{band}$ from $bandLabeledStorylinesRDD_{band}$ using flatMap, filter and distinct transforms
 5: {step 1: build LabeledData objects for each band}
 6: RDD<VectorAndLabeledData> $labeledDataRDD_{band}$ from $bandLabeledStorylinesRDD_{band}$ and $entityIndexRDD_{band}$ using map and filter transforms
 7: {step 2: build model for each band}
 8: $model_{band} = SVMWithSGD.train(labeledDataRDD_{band}, numIterations)$;
 9: {step 3: store training data term indices in dictionary}
 10: dictBandRDD = $entityIndexRDD_{band}$ training data terms with indexes;
 11: Store model and dictionary for the band
 12: end for

Algorithms used to build N classifiers for the N perception bands of peer brands is described in Algorithm 1. For each of the bands, training data provided by analysts consisting of labeled storylines is used to generate a String RDD of storylines and indexes from an integer indexed keywords dictionary of entities in the storylines in Step 0. Map transform operates on each element of an RDD in parallel and transforms it into another RDD of same length. FlatMap flattens RDD of N collections into a flat RDD of length N . PairRDD here represent an RDD of < Key, Value> tuples. The storyline RDD and index RDD is used to build an RDD of index vectors and LabeledData objects in Step 1. The classifiers are then built with the index vectors for each band's training dataset and the N models generated for the category for each band in Step 2 using MLLib's linear SVM library (SVMWithSGD). Finally the classifier and index RDDs are saved for the band for subsequent scoring of unlabeled storylines in Step 3.

Apply perception model

Algorithm 2 Generate perception using models for bands

Input: $\{storyline_i\}, \{bands\}, svmModel_{band}$ {storylines for new brand that needs its perception calculated and models for each band}
Output: $\{score_i, forbrand_i$ along with resulting perception band for brand}
 {calculate brand perception}
 1: {step 0: load testing data, dictionaries and models}
 2: RDD<String> $testStorylinesRDD$ using textfile method to read testing data for band on disk
 3: load $model_{band}$ and PairRDD<String, Long> $entityIndexRDD_{band}$
 4: for all band \in bands do
 5: {step 1: index testing data entities with training data for band}
 6: RDD<VectorAndUnLabeledData> $unlabeledDataRDD_{band}$ from $testStorylinesRDD$ and $entityIndexRDD_{band}$ using map and filter transforms
 7: broadcast $model_{band}$
 8: {step 2: score testing data with band model}
 9: PairRDD<StoryLinesResult> $scoredStoryLinesRDD_{band}$ using map transform by applying $model_{band}$ to each storyline in $unlabeledDataRDD_{band}$
 10: end for
 11: {step 3: Count positive scores for each band}
 12: for all band \in bands do
 13: PairRDD<StoryLinesResult> $posScoredStoryLinesRDD_{band}$ using map and filter transform on PairRDD<StoryLinesResult> $scoredStoryLinesRDD_{band}$
 14: S_j as count elements in $posScoredStoryLinesRDD_{band}$ using count action
 15: end for
 16: {step 4: calculate comprehensive measure with band positive counts}
 17: apply formula to calculate comprehensive score C_S

The algorithm used in DERIV to generate comprehensive measure using the final model built with the N SVM classifiers is described in Algorithm 2. Testing data consisting of storylines of test brand is generated as RDD of storylines in Step 0 along with loading the models for each band generated during training. Indexed vector is built using dictionary for each band's training data as RDD of entity and its index in Step 1. The models are broadcast to each of the nodes of the cluster so they can be applied to each of the storylines. The test data vectors are then scored against each model in Step 2 to generate an RDD of each storyline and its scores against each model. The counts of storylines for each band scoring above a threshold is calculated by applying filter transforms on RDD. Finally the final score is calculated by applying the comprehensive measure formula in Step 4.

These techniques show the effectiveness of using in-memory distributed techniques for calculating brand perception. The scoring of storylines is inherently parallelizable and is performed by broadcasting the models to each of the worker nodes in the cluster and using it to score storylines in parallel.

Experiments

This section presents experiments performed to show the effectiveness and scalability of DERIV brand perception tracking framework. They are implemented in Apache Spark in Java and run on AWS clusters. “[Experiment design](#)” subsection provides details of the datasets used and brands evaluated. “[Performance details](#)” subsection performance of the system in summarizing large number of storylines into perception and “[Case studies](#)” subsection describes the results and analysis of the measures for each brand tested. “[Analysis of results](#)” subsection analyses the use cases and performance results.

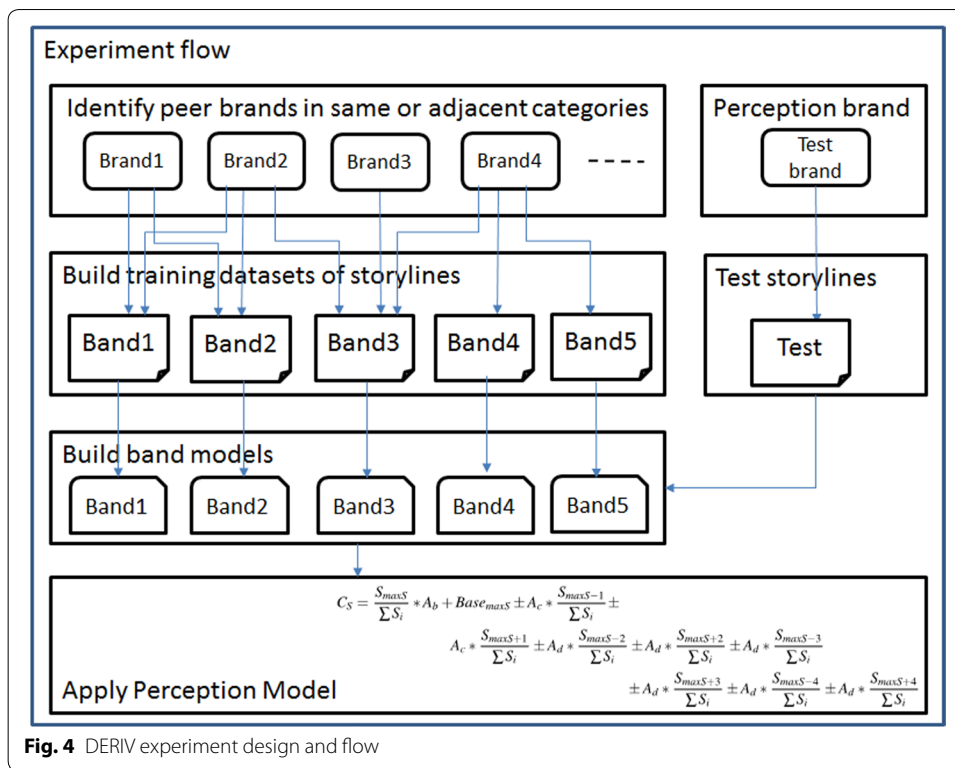
Experiment design

The experiment design flow is shown in Fig. 4. It shows peer brands for a category identified by an analyst. The training data is then built for each band using storylines from microblog text for the brands designated or adjacent band. The models built with training data is then used to score the storylines of the brand whose perception is measured. The counts of storylines above a threshold is used to model the perception of the brand.

Datasets

We performed experiments with three distinct datasets consisting of tweets to build the perception measure of three different brands in distinct categories. Tweets were collected in September and October of 2015 for the first two datasets, and in October and November of 2016 for the third dataset. These tweets were used to generate training storylines to build models and to score storylines for brands being measured against models for each band.

The first set of data consisted of tweets related to fashion apparel brands. Five apparel brands were selected, each representing one of the five bands of user perception previously defined. Tweets were collected using keywords related to each of the peer brands including the brands' name, stock symbol, terms associated with fashion apparel (for example, purse, heels, skirt, etc). The collected tweets were then used to generate storylines. Analysts labeled the resulting storylines as positively or negatively associated with the brand's pre-defined perception band. For instance, for deteriorating brands,



storylines generated from Tweets expressing lagging sales, increasing competition, poor customer service, or containing a negative tonality towards the product or company were labeled as positively associated with the declining brand. Conversely, for strengthening brands, storylines generated from Tweets expressing increasing sales, positive company news or containing a positive tonality towards the brand were labeled as positively associated with a strengthening brand. These labeled storylines were used as training data to build models to score storylines of a sixth fashion brand (referred to as Brand X) in order to score Brand X’s perception. The second dataset was on political candidates for a presidential election, while the third and final dataset was on the topic of electric vehicles. Based on five known brands for each respective topic, the perception of a sixth brand (Candidate Y for the presidential election and Car Model Z for electric vehicles) was generated through a process similar to the one described for fashion apparel.

SVM classifiers were built with labeled storylines for each band. The number of storylines in training data for each of the use cases in corresponding bands is shown in Table 1. The positive and negative storylines for each band were added to training data by analyst. It was crucial for analyst to mark storylines that indicated the peer brand belonging to a perception band as positive and a storyline irrelevant to the brand being in that band as negative. The models built with labeled storylines for each band are used to filter the large volumes of test storylines that can be potentially generated on the chosen brand and its domain in a short period and limit brand perception calculation to the ones that score highly against one or more models.

In order to aid the analyst in training data generation, several tools were provided to them. Not only could they search through storylines by keywords and their

Table 1 Number of storylines in each band for training for the use cases described

	Rapidly improving		Slowly improving		Holding steady		Slowly deteriorating		Rapidly deteriorating	
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
Fashion apparel	483	374	172	111	191	102	322	141	523	248
Political candidate	743	429	323	165	238	218	297	224	529	381
Electric car	181	144	177	145	122	201	222	147	143	188

combinations, but also visualize the graph of storylines to find other terms associated with the tweets from which those terms are extracted. The analysts could also search through tweets and identify the ones of interest and analyze the storylines generated from them. This helps analysts narrow down the storylines on peer brands that need to be included in training datasets for each band.

Methods

We compared the performance of classifiers based DERIV perception modeling with multi-class logistic regression and Sentiment Analysis based perception. The multi-class regression model was built with the positively labeled training data elements for each band. Linear SVM classifiers with L2 regularization were found to be most accurate and used in model building. Logistic Regression with Stochastic Average Gradient (SAG) multinomial solver and L2 penalty was used [31]. Sentiment analysis was performed on tweets from which the storylines scoring above threshold for each band were generated for comparison using Stanford Core NLP [32].

Performance

The performance of the perception modeling and performance of distributed algorithms is now described.

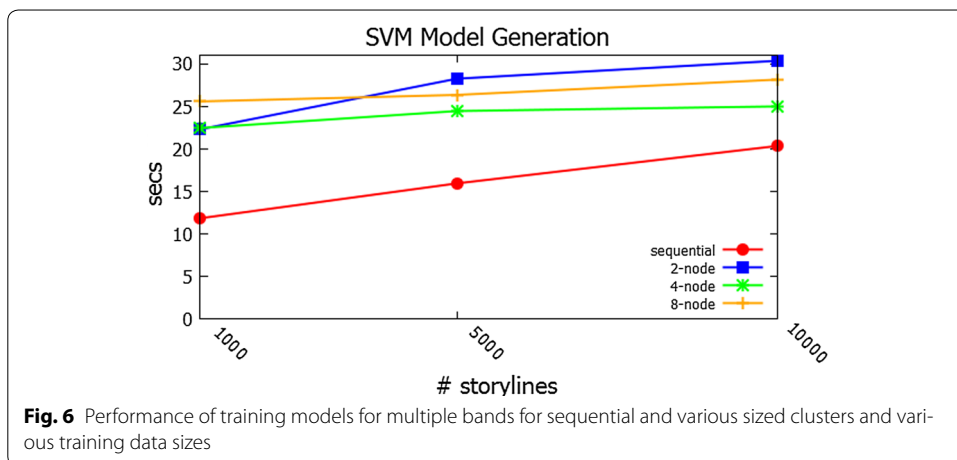
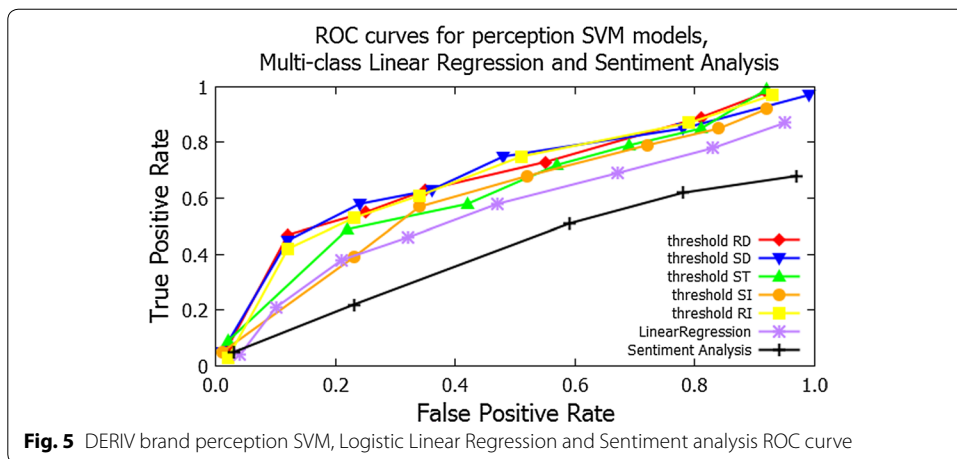
Qualitative effectiveness

To validate the *accuracy* of perception measure, different metrics were adopted: the True Positive Ratio (TPR) designates the percentage of perception designations that successfully matched the perception as specified by analyst as true, while the False Positive Ratio (FPR) denotes the percentage of perception designations that were actually incorrect. In addition, a ROC curve was utilized to evaluate the perception performance as its discrimination threshold for each predictive model was varied. The values of the enumerated labels for positive, negative and neutral sentiment were varied for sentiment analysis. The graphs of the ROC curves is shown in Fig. 5. Since the sentiment analysis model was trained on corpus of long documents, its performance on short text of tweets was poor. Multi-class logistic regression also did not fare as well. SVM models for each of the bands performed the best. Sentiment analysis can however be useful to highlight storylines from highly positive or negative sentiment tweets as possible candidates for positively labeled training data for deteriorating or improving perception bands to reduce analyst's workload.

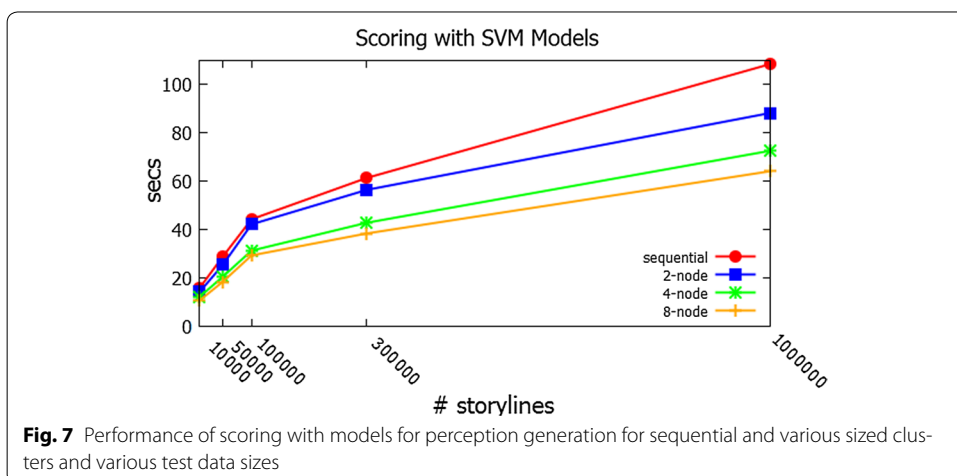
Quantitative effectiveness

The *computational* performance of the techniques used in models creation and scoring for perception calculation at different levels of distribution is evaluated in this subsection. The results for running the techniques on various sized clusters and dataset sizes are presented. For sequential or single node experiments, a MacBook Pro with 16 GB RAM and a 4 core 2.5 GHz Intel i7 processor was used. For cluster experiments, Amazon EC2 instances of type m3.2x large with 8 vCPUs and 32GB RAM were used for master and slaves.

In Fig. 6, the times for building the SVM models with the multiple bands and sizes of training data is shown. It clearly shows the improvement in time with increasing sized



clusters. However building the models on a single node setup is faster for small enough data sets while on larger clusters it is higher initially but does not increase significantly for increasing data sizes. In Fig. 7 the improvement in performance of scoring over



larger sized clusters on increasing data sizes is presented. As it becomes difficult to score larger datasets on a single node the scaling on spark cluster can continue horizontally indefinitely by adding nodes to clusters.

The calculation of the perception measure is extremely efficient and scales well with increasing number of scored storylines as shown in Fig. 8 showing the performant nature of the perception calculation. The root cause analysis performance is shown in Fig. 9. The figure shows the ease with which the original tweets for the storyline entities can be mapped back to them.

Case studies

The perception measures and the underlying storylines were very revealing for each of the three brands in three different categories for which perception modeling were performed.

Fashion apparel

Storylines for fashion apparel brand were scored against the models, the largest number of storylines had positive scores above threshold for the RD band, followed closely by

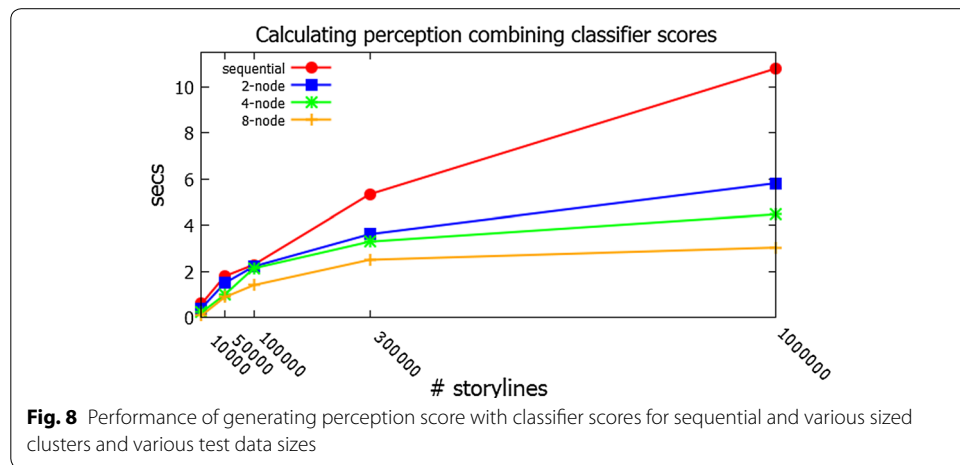


Fig. 8 Performance of generating perception score with classifier scores for sequential and various sized clusters and various test data sizes

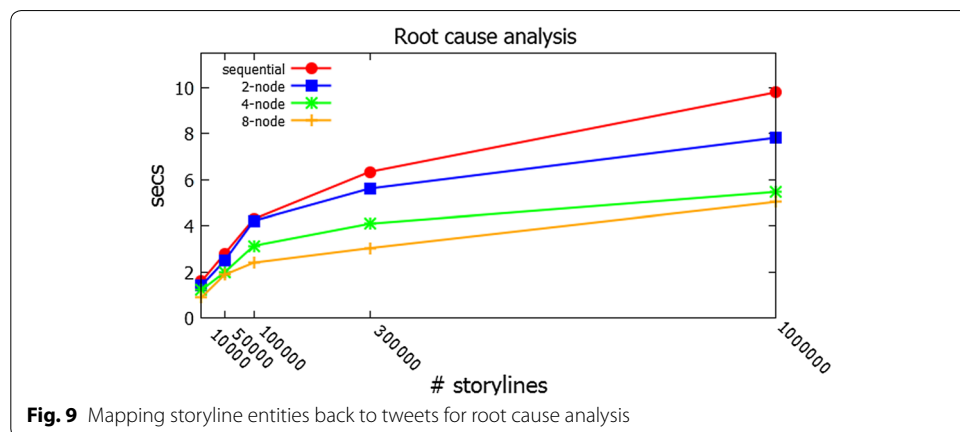


Fig. 9 Mapping storyline entities back to tweets for root cause analysis

the SD brand. Of the 19,336 scored storylines generated from 7898 tweets, 3097 were positively labeled as rapidly improving, 3207 as slowly improving, 3566 as stable, 5960 as slowly deteriorating and 6609 as rapidly deteriorating for SVM threshold set to 0.5. Based on our formula and calculations, the resulting brand perception score of apparel Brand X was 5.44. A sample of some of the top storylines with scores associating Brand X with a rapidly and slowly deteriorating perception are shown in Table 2. These storylines include the terms ‘men’, ‘bags’, ‘#deals’, ‘nike’. The brand labeled as strongly deteriorating in the training dataset experienced sales slumps during the experiment period in their line of purses and mens fashion, thus explaining the association of Brand X with ‘men’ and ‘bags’ as indicative of a declining brand. The storylines also indicated that Brand X was suffering from many of the issues afflicting other fashion brands that have recently struggled in a competitive retail environment filled with heavy discounting (#deals) and significant promotions necessitated by a strong U.S. dollar. Additionally, many high-end apparel brands, of which Brand X is one, have suffered from the societal move towards the acceptance of athleisure (Nike) as everyday wear, which has pressured sales for these higher end brands. Brand X’s suffering brand perception is further evidenced by revenues and earnings that missed Wall Street’s expectations and a stock price that saw a 25% decline in the three months preceding the date of the dataset.

The representative storylines with top scores in rapidly and slowly deteriorating bands are shown in Table 2.

Political candidate

Out of 7559 storylines scored for political candidate from 3066 tweets, 1687 were labeled rapidly deteriorating, 1696 were slightly deteriorating, 1537 as stable, 974 as slightly improving and 1365 as rapidly improving for SVM threshold set to 0.5. Based on our formulae the comprehensive brand perception for the presidential candidate was calculated to be 26.07. For the analysis of Presidential Candidate Y, a sample of several of the top scoring storylines for the slowly and rapidly deteriorating bands is shown in Table 3. The terms ‘women’, ‘feminist’, ‘liar’, ‘isis’ and ‘the establishment’ are terms that show up again and again for Candidate Y. This is indicative of voters’ backlash towards presidential candidates that are considered part of ‘the establishment’ and also show the public’s displeasure of Candidate Y’s proposed handling of ISIS. There have also been rampant accusations of Candidate Y’s spinning of the facts which have led many to accuse the candidate of being a liar. The perception of 26.07, which places Candidate Y in the slowly

Table 2 Storylines for fashion apparel with the highest band scores for the rapidly and slowly deteriorating bands

Storylines for rapidly deteriorating band

Fashion brand X → men → 5 → the best weekend bags

Fashion brand X → #deals → polo ralph lauren mens sneakers → 5

Fashion brand X → i actually → underwear → nike

Storylines for slowly deteriorating band

Fashion brand X → ralph lauren men:ralphlauren check → \$45

Fashion brand X → ralph lauren men size → neck style #sweater blue color → stripes

Fashion brand X → polo black → ralphlauren → stripes

deteriorating band is corroborated by the candidates decreasing poll numbers in the weeks after this dataset was produced. The top scoring storylines for rapidly and slowly deteriorating band are shown in Table 3.

Electric car

Of the 4499 storylines scored for electric car use case generated from 35,868 tweets, 1928 scored above the threshold for slowly improving band, 4460 scored above threshold for rapidly improving band, 1480 for stable band, 860 for slowly deteriorating band and 1491 for rapidly deteriorating band. This gives the perception score for the brand as 90.1.

A sample of the top scoring storylines for the rapidly improving and slowly improving bands is depicted in Table 4. ‘MLCC capacitors target ev,’ ‘station,’ ‘charging,’ ‘collaborate to build,’ and ‘infrastructure’ are terms that appear in the rapidly improving band, while ‘#solarnetworks,’ ‘kw super-fast ccs electric,’ ‘100 bay area i3 owners’ appear in the slowly improving band. These terms mostly relate to the charging infrastructure for electric vehicles. Two of Car Model Z’s strengths include its miles per gallon equivalent rating as well the efficient charging time of its battery, which could get to 80% charging in just 20 minutes using a Fast Charging station. Other terms in the rapidly improving band and slowly improving band include ‘Daimler,’ ‘collaborate to build,’ ‘@daimler plan eu fast’ all in reference to the collaboration of automakers in building out the car charging

Table 3 Storylines for political candidate with the highest score for slowly and rapidly deterioration perception bands

Storylines for slowly deteriorating band
Candidate Y → women → the establishment → the difference
Candidate Y → a staunch feminist → a liar → gop
Candidate Y → isis → existence right now → the establishment
Candidate Y → women → the establishment → more rino gop too

Storylines for rapidly deteriorating band
Candidate Y → jeb bush → rights → blacks
Candidate Y → koch brothers favorites → jeb bush → t. boone pickens
Candidate Y → flagrant liar → isis → hewlett
Candidate Y → isis → climate change → our stafford republican women

Table 4 Storylines for political candidate with the highest score for slowly and rapidly deterioration perception bands

Storylines for rapidly improving band
Car Brand Z → electric vehicle → station → charging
Car Brand Z → daimler → ev → collaborate to build
Car Brand Z → electric vehicle → infrastructure → discover
Car Brand Z → electric vehicle → ev → mlcc capacitors target ev

Storylines for slowly improving band
Car Brand Z → electric vehicle → bmw → @daimler plan eu fast
Car Brand Z → bmw → europe → #solarnetworks
Car Brand Z → electric vehicle → bmw → 100 bay area i3 owners
Car Brand Z → electric vehicle → bmw → kw super-fast ccs electric

infrastructure throughout the U.S. and Europe. The high brand perception score is also supported by the generally favorable reviews in the media and from consumers.

Analysis of results

The qualitative performance results show the DERIV framework's ability to distill storylines into a perception that is more accurate than sentiment analysis scores or multi-class logistic regression models as shown by AUC measure. The quantitative performance results show the ability to scale the supervised model building, scoring new storylines and perception calculations to a large number of storylines to be able to measure perception as frequently as desired. All this is possible primarily due to the ability to leverage in-memory distribution techniques offered in widely available open source framework. Performing the processing on public cloud allows for horizontal scalability on demand if data suddenly increases.

The use cases depict the effectiveness of the scores in measuring brand perception in diverse fields of electric cars, fashion and political candidates from the storylines of brand being measured with models built with storylines of peer brands in respective categories. The perception validates analyst's view of what is happening with the perception of brand at the time demonstrating the meaningfulness of the measure generated entirely from customer voice in social data.

The separation of storylines that score above the threshold is crucial for the formulae in "[Comprehensive perception modeling](#)" section. The labeling of the storylines with as little overlap as possible in the positive and negative label sets is key to that. This makes the labeling exercise a crucial and labor intensive activity for the analyst. The analyst can be aided by the presentation of storylines from highly positive and negative tweets as determined by sentiment analysis. Analysts are also helped by storylines that score just below the threshold against the band models as candidates for training set.

Conclusions

Brand perception measurement through sentiment analysis is often inaccurate and surveys are also archaic. Our technique interprets customer voice from social media and other open data by connecting the dots across data elements as storylines and using them to measure brand perception. It calculates perception based on peer brands storylines labeled for various bands of perception and supervised learning models built from them. A combination of machine learning and statistical techniques provide a highly effective and accurate way to measure perception and its changes. Distributed in-memory algorithms allow computing perception at scale by including all relevant customer voice sources and scaling to the large number of storylines from sources like Twitter. Extensive experiments with multiple brands from diverse categories validate perception distilled from storylines as effective in capturing true customer voice.

Abbreviations

DISCRN: Distributed Spatio-temporal Conceptsearch based Storytelling; DERIV: Distributed, in-memory framework for tracking consumer voice; SVM: Support Vector Machine; RI: rapidly improving; SI: slowly improving; S: steady; SD: slowly deteriorating; RD: rapidly deteriorating; AWS: Amazon Web Services; EC2: Elastic Compute Cloud.

Authors' contributions

Manu Shukla was the primary author with Raimundo Dos Santos, Andrew Fong and Chang-Tien Lu contributing with refining the concepts discussed and format of the paper. All authors read and approved the final manuscript.

Author details

¹ Omniscience Corp, Palo Alto, CA, USA. ² US Army Corps of Engineers-GRL, Alexandria, VA, USA. ³ Virginia Tech, Falls Church, VA, USA.

Acknowledgements

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Availability of data and supporting materials

The data and material for paper are not being made available.

Consent for publication

All authors have consented for publication of this paper.

Ethics approval and consent to participate

All authors give ethics approval and consent to participate in submission and review process.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 29 March 2017 Accepted: 5 June 2017

Published online: 17 June 2017

References

1. Wiersma W. The validity of surveys: online and offline. 2013.
2. Kervyn N, Fiske ST, Malone C. Research dialogue. *J Consum Psychol*. 2012;22(2):166–76.
3. Ahmad FS, Ihtiyar A, Jing W, Osman M. Integrating brand perception, culture dimension and social influence in predicting purchase intention in luxury brand market. In: Third international conference on business and economic research, Indonesia. 2012.
4. Hanzae KH, Rouhani FR. Investigation of the effects of luxury brand perception and brand preference on purchase intention of luxury products. *Afr J Bus Manag*. 2013;7(18):1778–90.
5. Clemenz J, Brettel M, Moeller T. How the personality of a brand impacts the perception of different dimensions of quality. *J Brand Manag*. 2012;20(1):52–64.
6. Lindstrom M. *Brand sense: sensory secrets behind the stuff we buy*. Simon and Schuster. 2008.
7. Hardie BG, Johnson EJ, Fader PS. Modeling loss aversion and reference dependence effects on brand choice. *Mark Sci*. 1993;12(4):378–94.
8. Fader PS, Schmittlein DC. Excess behavioral loyalty for high-share brands: deviations from the dirichlet model for repeat purchasing. *J Mark Res*. 1993;30(4):478–93.
9. van Rompay TJL, Pruyn ATH. When visual product features speak the same language: effects of shape-typeface congruence on brand perception and price expectations*. *J Prod Innov Manag*. 2011;28(4):599–610.
10. Rafique M. Impact of celebrity advertisement on customers' brand perception and purchase intention. *Asian J Bus Manag Sci*. 2012;1(11):53–67.
11. Ziegler CN, Skubacz M. Towards automated reputation and brand monitoring on the web. In: 2006 IEEE/WIC/ACM international conference on web intelligence (WI 2006 main conference proceedings)(WI'06). New Jersey: IEEE; 2006. p. 1066–72.
12. Cambria E, Schuller B, Xia Y, Havasi C. New avenues in opinion mining and sentiment analysis. *IEEE Intell Syst*. 2013;28(2):15–21.
13. Erdmann M, Ikeda K, Ishizaki H, Hattori G, Takishima Y. Feature based sentiment analysis of tweets in multiple languages. In: Benatallah B, Bestavros A, Manolopoulos Y, Vakali A, Zhang Y, editors. *Web Information Systems Engineering – WISE 2014, volume 8787 of Lecture Notes in Computer Science*. Berlin: Springer International Publishing; 2014. p. 109–124.
14. Ghiassi M, Skinner J, Zimbra D. Twitter brand sentiment analysis: a hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Syst Appl*. 2013;40(16):6266–82.
15. Mostafa MM. More than words: social networks' text mining for consumer brand sentiments. *Expert Syst Appl*. 2013;40(10):4241–51.
16. Chamlerwat W, Bhattarakosol P, Rungkasiri T, Haruechaiyasak C. Discovering consumer insight from twitter via sentiment analysis. *J UCS*. 2012;18(8):973–92.
17. da Silva NF, Hruschka ER. Tweet sentiment analysis with classifier ensembles. *Decis Support Syst*. 2014;66:170–9.
18. Sidorov G, Miranda-Jiménez S, Viveros-Jiménez F, Gelbukh A, Castro-Sánchez N, Velásquez F, Díaz-Rangel I, Suárez-Guerra S, Treviño A, Gordon J. Empirical study of machine learning based approach for opinion mining in tweets. In: Mexican international conference on artificial intelligence. Berlin: Springer; 2012. p. 1–14.
19. Gundecha P, Liu H. Mining social media: a brief introduction, Chapter 2, 2012. p. 1–17.

20. Kondreddi S, Triantafillou P, Weikum G. Combining information extraction and human computing for crowdsourced knowledge acquisition. In: IEEE 30th international conference on data engineering (ICDE). 2014. p. 988–99.
21. Tang S, Wu F, Li S, Lu W, Zhang Z, Zhuang Y. Sketch the storyline with charcoal: a non-parametric approach. In: Proceedings of the 24th international conference on artificial intelligence. Quebec : AAAI Press; 2015. p. 3841–8.
22. Hu P, Huang M-L, Zhu X-Y. Exploring the interactions of storylines from informative news events. *J Comput Sci Technol*. 2014;29(3):502–18.
23. Wang D, Tao Li MO. Generating pictorial storylines via minimum-weight connected dominating set approximation in multi-view graphs. In: Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence. 2014. pp. 683–9.
24. Dos Santos RF, Shah S, Chen F, Boedihardjo A, Butler P, Lu CT, Ramakrishnan N. Spatio-temporal storytelling on twitter. In: Virginia tech computer science technical report. 2015. <http://vtechworks.lib.vt.edu/handle/10919/24701>.
25. Langner T, Bruns D, Fischer A, Rossiter JR. Falling in love with brands: a dynamic analysis of the trajectories of brand love. *Mark Lett*. 2016;27(1):15–26.
26. Zaharia M, Chowdhury M, Das T, Dave A, Ma J, McCauly M, Franklin MJ, Shenker S, Stoica I. Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing. In: Presented as part of the 9th USENIX symposium on networked systems design and implementation (NSDI 12). San Jose: USENIX; 2012. p. 15–28.
27. Shukla M, Santos RD, Chen F, Lu CT. DISCRN: a distributed storytelling framework for intelligence analysis. Virginia Tech Computer Science Technical Report. <http://hdl.handle.net/10919/53944>. 2015.
28. Santos RFD, Shah S, Boedihardjo A, Chen F, Lu C-T, Butler P, Ramakrishnan N. A framework for intelligence analysis using spatio-temporal storytelling. *Geoinformatica*. 2016;20(2):285–326.
29. Apache and Spark. <http://spark.apache.org>. 2015.
30. MLlib and SVM. <https://spark.apache.org/docs/latest/mllib-linear-methods.html#linear-support-vector-machines-svms>. 2015.
31. scikit-learn and LR. http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html. 2016.
32. Manning CD, Surdeanu M, Bauer J, Finkel J, Bethard SJ, McClosky D. The Stanford CoreNLP natural language processing toolkit. In: Proceedings of 52nd Annual Meeting of the Association for computational linguistics (ACL) system demonstrations. Baltimore, Maryland, USA; 2014. p. 55–60.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
