DERIV: Distributed In-memory Brand Perception Tracking Framework

Manu Shukla and Andrew Fong
Omniscience Corporation
Palo Alto, CA

Raimundo Dos Santos
US Army Corps of Engineers - GRL
Alexandria, VA

Chang-Tien Lu
Computer Science Dept., Virginia Tech
Falls Church, VA

***Abstract*—Social media captures voice of customers at a rapid pace. Consumer perception of a brand is crucial to its success. Current techniques for measuring brand perception using lengthy surveys of handpicked users in person, by mail, phone or online are time consuming and increasingly inadequate. A more effective technique to measure brand perception is to interpret customer voice directly from social media and other open data. In this work we present DERIV, a DistributEd, in-memoRy framework for trackIng consumer Voice based on a brand perception measure using storylines generated from open data. The framework measures 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 accurate and effective in capturing brand perception.**

I. Introduction

Tracking brand perception, both online and offline, has been ill-served due to time-consuming techniques used by survey companies [Wiersma, 2013]. Online survey companies take pre-set questions from client companies and present them to users. Offline survey companies handpick representative users and ask them detailed questions about products. Responses are then carefully analyzed making the process time consuming and expensive. These methods are not only antiquated but also require a long time to execute. 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. It utilizes *storylines* generated from open data using the DISCRN framework [Shukla et al., 2015]. DISCRN is a distributed framework that connects entities across open data elements with relationships as storylines. Each entity in a storyline is extracted from a data element and is connected to other entities from the same or subsequent data element through relationships or verbs that precede the entity in that data element. Storylines are built with open data such as tweets about the brand with entities extracted from data elements used in the generation of *ConceptGraph* [Santos et al., 2016]. The ConceptGraph is subsequently traversed using *ConceptSearch* to generate the storylines which are a more powerful instrument to capture user voice than any user sentiment measure. The key reasons for measuring user perception from storylines are:

1) **Connect 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, generated by connecting entities across sources and individual elements, offer a comprehensive view of user's perception of a brand.
2) **Compact representation of customer voice:** The storylines are a compact representation of a set of tweets and posts whose entities and relationships are connected in meaningful ways.
3) **Eliminate noise:** Connecting entities across a source such as tweets and modeling with storylines eliminates the clutter of brand specific terms and verbiage that is prevalent in tweets.

As an example of how storylines are generated from a source such as tweets, 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 y $\rightarrow$ new #koch brothers darling! $\rightarrow$ #support the #establishment $\rightarrow$ #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.

The key contributions of the paper are:

1) **Design algorithms to measure brand perception from storylines utilizing supervised learning**: Created novel algorithms to measure perception based on supervised learning models from training data generated by analysts. The models are used to distill a comprehensive perception of a brand.
2) **Develop framework based on in-memory distributed techniques to perform supervised model building and scoring at scale**: The algorithms in DERIV use distributed in-memory techniques that scale the building of multiple models with increasing number of labeled storylines. The scaling also allows scoring large number of new storylines about a brand every time period and generate comprehensive brand perception.
3) **Conduct extensive experiments validating perception from open data**: Perform experiments with Twitter data

978-1-5090-6167-9/16 $31.00 © 2016 IEEE
DOI 10.1109/ICMLA.2016.77

on several brands in diverse domains. The results show the relevance and effectiveness of calculated perception as compared to sentiment analysis.

II. Related Works

Marketing oriented perception measures have been researched for a long time. Social perception theory has been used to measure brand perception [Kervyn et al., 2012]. Cultural dimension and social influence on brand perception is examined [Ahmad et al., 2012]. Impact of brand perception on luxury item purchases is explored [Hanzaee and Rouhani, 2013]. Connection between quality and perception of a brand has been investigated [Clemenz et al., 2012]. Users sense of a brand has also been studied [Lindstrom, 2008]. Users selection of a brand based on multiple factors is explored [Hardie et al., 1993]. Experiments with high-share brands loyalty is described [Fader and Schmittlein, 1993]. Effect of shape on brand perception is discussed [van Rompay and Pruyn, 2011]. Impact of celebrity on brand perception is investigated [Rafique, 2012]. Sentiment analysis has been used extensively to measure user sentiment towards a brand. Extracting sentiments from tweets has been explored [Erdmann et al., 2014] while brand sentiment analysis has also been studied [Ghiassi et al., 2013]. Detecting polarity in tweets helps in gauging customer sentiment towards a brand [Chamlertwat et al., 2012]. Classifier ensembles have been explored for tweet sentiment analysis [da Silva et al., 2014]. The sentiment analysis techniques however focus on individual tweets and have no ability to detect perceptions by connecting sets of multiple tweets.

Social media mining has been a rich source of information on brands [Gundecha and Liu, 2012]. Use of social media for knowledge acquisition and validation is well known [Kondreddi et al., 2014]. Linking new articles together to generate evolving new stories is popular [Tang et al., 2015]. Interactions of storylines in news is explored [Hu et al., 2014]. Building storylines of text, pictorial and structured data is investigated [Dingding Wang, 2014]. Storylines have been used to determine evolving events effectively [Dos Santos Jr et al., 2015]. Storylines to the best of our knowledge have never been used to measure brand perception.

III. DERIV Brand Perception

This section describes the modeling and scoring techniques used to generate the DERIV perception. Section III-A provides overview of steps in perception calculation, bands, modeling and scoring used in brand perception. Section III-B details the measure calculation.

A. DERIV Flow

DERIV employs storylines and models built with supervised learning techniques to generate perception. Figure 1 shows the flow of data and operations in DERIV. The first block in the figure shows storylines being processed in parallel to

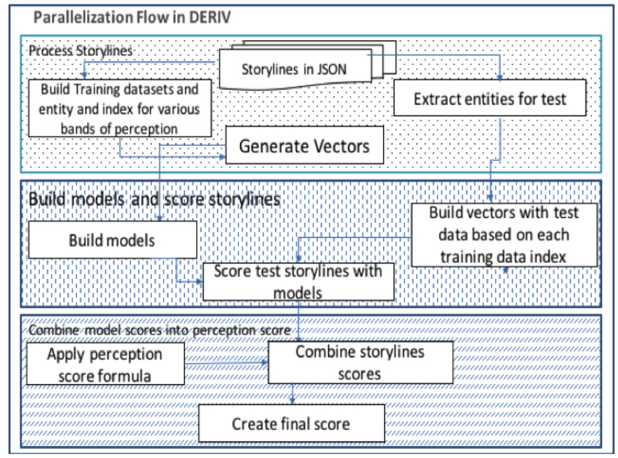

Fig. 1. Jobs and transforms needed to calculate perception.

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 in each band are counted and their counts used to calculate comprehensive perception in final block. The bands are pre-determined slices of perceptions in which a known peer brand is determined to be situated based on pre-existing survey data or industry standard measures.

An analyst determines filtering keywords to collect open data for a brand to be tracked and each of its peer brands. Data is subsequently used to build storylines from which an analyst labels relevant ones as representative or non-representative to that brand's perception within the band. Some examples of perception bands are rapidly improving, stable, slowly deteriorating etc. 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 together to constitute a model. Linear SVM (Support Vector Machines) classifiers with L2 regularization were found to be most accurate and used in model building. The combination of 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 determines the band perception will lie in and the counts for other bands are used to tweak perception further within the band.

B. Comprehensive Perception

The perception is based on the number of positively labeled storylines in a time period and represent one of the N bands. The bands used in this study are:

1) Rapidly improving (RI) - Models rapidly improving brand perception.
2) Slowly improving (SI) - Represents slowly improving brand perception.
3) Holding steady (S) - Models stable brand perception.
4) Slowly deteriorating (SD) - Represents slowly deteriorating brand perception.
5) Rapidly deteriorating (RD) - Models rapidly deteriorating brand perception.

The classifiers for each of the bands above give a score to the storyline. The final perception is then calculated using the formula in Equation 1 when the maximum number of storylines score above threshold δ for rapidly improving band.

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

When the maximum number of storylines scoring above the threshold are in slowly improving band the final measure is calculated using the formula in Equation 2.

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

The final perception is calculated using the formula in Equation 3 when the maximum count is for the storylines that score in stable band.

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

When the maximum count is for the storylines scoring in slowly deteriorating band the final measure is calculated using the formula in Equation 4.

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

When the maximum count is for the storylines scoring in rapidly deteriorating band the final perception is calculated using the formula in Equation 5.

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

In equations 1-5 S_i represents the count of positive scores of the storylines for a given model S. The final perception equation can be written in a generalized form as depicted in Equation 6.

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

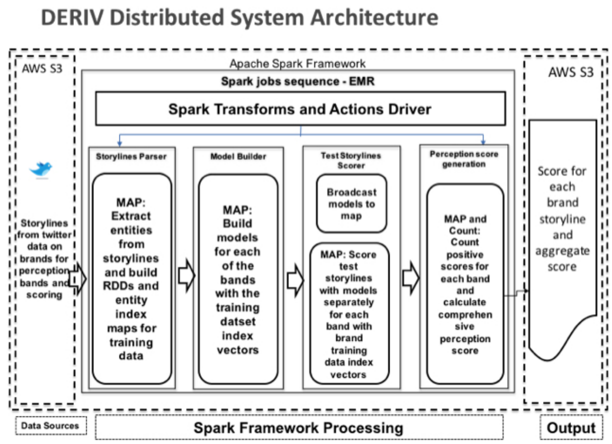

Fig. 2. DERIV brand perception calculation framework architecture.

C_S represents the cumulative DERIV perception. A_i represents the weights assigned to each band's positively labeled count of storylines. Default A_b is 20 for N as 5, A_c is 10 and A_d is 5. The values assigned to the A_i coefficients depend on their adjacency to the band with the highest count. Default $Base_{RI}$ is 80, $Base_S$ is 40, $Base_{SD}$ is 20 and $Base_{RD}$ is 0. 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. For example, if a brand's open data elements generate 100 storylines that score above threshold for band models, 25 of which score above the threshold for slowly improving, 45 for rapidly improving, 20 for stable and 10 for slowly deteriorating band models, 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 equation has negative sign for S_{RD} and S_{SD} for the coefficients A_i when i $\in$ {c,d} and positive otherwise.

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.

IV. DERIV SYSTEM

This section presents the architecture of the DERIV framework and provides the details of distributed in-memory algorithms. Section IV-A describes the architecture of the system and Section IV-B describes the algorithms used.

A. Architecture

The architecture of DERIV is shown in Figure 2. The DERIV framework is a sequence of Spark jobs [Zaharia et al., 2012] that run on AWS EC2 (Elastic Compute Cloud) clusters and continues with storylines generated by DISCRN. It proceeds to generate 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 [Apache and Spark, 2015]. The SVM classifiers used are from Spark MLLib library [MLLib and SVM, 2015]. The programming constructs available in Spark are reading of data on disk into RDDs (Resilient Distributed Datasets) 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 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 models with training data RDDs if new training data is provided. The third module iteratively scores testing data against each band's model using the dictionary indexes for training data and keeps scores in storylinesResults object's RDD. The fourth module generates counts of positive scores for each model that are above the threshold and applies formula for calculating the comprehensive DERIV measure.

B. Algorithms

The algorithm used in DERIV to generate the *N* models, one for each band for known peer brands and subsequently to generate perception measure with test storylines from the band models are described in this section. Algorithms used to build *N* models 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 models and index RDDs are saved for the band for subsequent scoring of unlabeled storylines in Step 3.

Algorithm 1 Generate models from training data

Input: $\{storyline_i\}$,$\{band_k\}$,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}
{step 0: build vectors and index for each band labeled data}
for all band $\in$ bands **do**
RDD<String>$bandLabeledStorylinesRDD_{band}$ using textfile method to read labeled training data for band on disk
PairRDD<String, Long>$entityIndexRDD_{band}$ from $bandLabeledStorylinesRDD_{band}$ using flatMap, filter and distinct transforms
{step 1: build LabeledData objects for each band}
RDD<VectorAndLabeledData>$labeledDataRDD_{band}$ from $bandLabeledStorylinesRDD_{band}$ and $entityIndexRDD_{band}$ using map and filter transforms
{step 2: build model for each band}
$model_{band}$ = SVMWithSGD.train($labeledDataRDD_{band}$, numIterations);
{step 3: store training data term indices in dictionary}
dictBandRDD = $entityIndexRDD_{band}$ training data terms with indexes;
Store model and dictionary for the band
end for

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}
{step 0: load testing data, dictionaries and models}
RDD<String>$testStorylinesRDD$ using textfile method to read testing data for band on disk
load $model_{band}$ and PairRDD<String,Long>$entityIndexRDD_{band}$
for all band $\in$ bands **do**
{step 1: index testing data entities with training data for band}
RDD<VectorAndUnLabeledData>$unlabeledDataRDD_{band}$ from $testStorylinesRDD$ and $entityIndexRDD_{band}$ using map and filter transforms
broadcast $model_{band}$
{step 2: score testing data with band model}
PairRDD<StoryLinesResult>$scoredStoryLinesRDD_{band}$ using map transform by applying $model_{band}$ to each storyline in $unlabeledDataRDD_{band}$
end for
{step 3: Count positive scores for each band}
for all band $\in$ bands **do**
PairRDD<StoryLinesResult>$posScoredStoryLinesRDD_{band}$ using map and filter transform on PairRDD<StoryLinesResult>$scoredStoryLinesRDD_{band}$
S_I as count elements in $posScoredStoryLinesRDD_{band}$ using count action
end for
{step 4: calculate comprehensive measure with band positive counts}
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 in Step 0. Indexed vector is built using dictionary for each band's training data in Step 1. The vectors are then scored against each model in Step 2. The final score is then calculated in Step 3 by counting positive scores for each band and 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.

V. 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. Subsection V-A provides details of the datasets used and brands evaluated. Subsection V-B describes the results and analysis of the measures for each brand tested and Subsection V-C details performance of the system in summarizing large number of storylines into perception.

A. Experiment Design

We performed experiments with two distinct datasets consisting of tweets to build the perception measure of two different brands in distinct domains. Tweets were collected in September and October of 2015. Training data was generated by analysts and SVM classifiers built with them for each band had accuracy of 80% or higher from area under ROC (Receiver Operating Characteristic). Sentiment analysis was also performed on tweets from which the storylines scoring above threshold for each band were generated for comparison using Stanford Core NLP [Manning et al., 2014].

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 predefined 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) whose perception needed to be calculated. The second dataset was on political candidates for a presidential election. Based on five known candidates, the perception of a sixth candidate (Candidate Y) was generated through a process similar to the one described for fashion apparel.

B. Experiment Results

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 the SD brand. Of the 19,336 scored storylines, 3,097 were positively labeled as rapidly improving, 3,207 as slowly improving, 3,566 as stable, 5,960 as slowly deteriorating and 6,609 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 I. 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 I.

storylines for rapidly deteriorating band
Brand X → men → 5 → the best weekend bags
Brand X → #deals → polo ralph lauren mens sneakers → 5
Brand X → i actually → underwear → nike
storylines for slowly deteriorating band
Brand X → ralph lauren men:ralphlauren check → $45
Brand X → ralph lauren men size → neck style #sweater blue color → stripes
Brand X → polo black → ralphlauren → stripes

TABLE I
STORYLINES FOR FASHION APPAREL WITH THE HIGHEST BAND SCORES FOR THE RAPIDLY AND SLOWLY DETERIORATING BANDS.

Out of 7,559 storylines scored for political candidate, 1,687 were labeled rapidly deteriorating, 1,696 were slightly deteriorating, 1,537 as stable, 974 as slightly improving and 1,365 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 II. 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 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

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	
storyline 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 II
STORYLINES FOR POLITICAL CANDIDATE WITH THE HIGHEST SCORE FOR SLOWLY AND RAPIDLY DETERIORATION PERCEPTION BANDS.

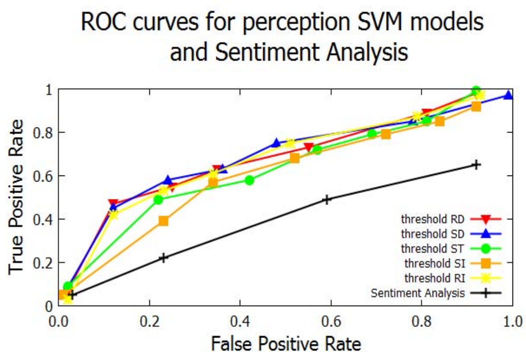

Fig. 3. DERIV brand perception SVM ROC curves and Sentiment analysis ROC curve.

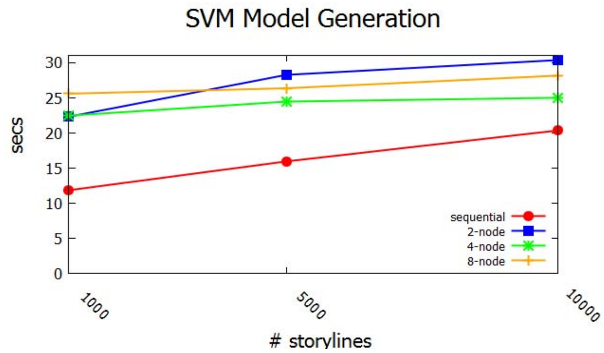

Fig. 4. Performance of training models for multiple bands for sequential and various sized clusters and various training data sizes

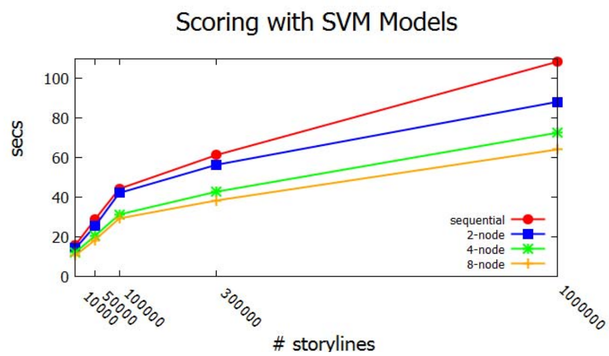

Fig. 5. Performance of scoring with models for perception generation for sequential and various sized clusters and various test data sizes

deteriorating band are shown in table II.

C. Performance

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 Figure 3. Since the sentiment analysis model was trained on corpus of long documents, its performance on short text of tweets was poor.

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 16GB RAM and a 4 core 2.5GHz Intel i7 processor was used. For cluster experiments, Amazon EC2 instances of type m3.2xlarge with 8 vCPUs and 32GB RAM were used for master and slaves. In Figure 4, 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 Figure 5 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.

VI. 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. The technique provides 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 for multiple brands validate perception distilled from storylines is effective in capturing true customer voice.

REFERENCES

[Ahmad et al., 2012] Ahmad, F. S., Ihtiyar, A., Jing, W., and Osman, M. H. M. (2012). 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*.

[Apache and Spark, 2015] Apache and Spark (2015). http://spark.apache.org.

[Chamlertwat et al., 2012] Chamlertwat, W., Bhattarakosol, P., Rungkasiri, T., and Haruechaiyasak, C. (2012). Discovering consumer insight from twitter via sentiment analysis. *J. UCS*, 18(8):973–992.

[Clemenz et al., 2012] Clemenz, J., Brettel, M., and Moeller, T. (2012). How the personality of a brand impacts the perception of different dimensions of quality. *Journal of Brand Management*, 20(1):52–64.

[da Silva et al., 2014] da Silva, N. F., Hruschka, E. R., and Jr., E. R. H. (2014). Tweet sentiment analysis with classifier ensembles. *Decision Support Systems*, 66:170 – 179.

[Dingding Wang, 2014] Dingding Wang, Tao Li, M. O. (2014). 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*, pages 683–689.

[Dos Santos Jr et al., 2015] Dos Santos Jr, R. F., Shah, S., Chen, F., Boedihardjo, A., Butler, P., Lu, C.-T., and Ramakrishnan, N. (2015). Spatio-temporal storytelling on twitter. *Virginia Tech Computer Science Technical Report http://vtechworks.lib.vt.edu/handle/10919/24701*.

[Erdmann et al., 2014] Erdmann, M., Ikeda, K., Ishizaki, H., Hattori, G., and Takishima, Y. (2014). Feature based sentiment analysis of tweets in multiple languages. In Benatallah, B., Bestavros, A., Manolopoulos, Y., Vakali, A., and Zhang, Y., editors, *Web Information Systems Engineering WISE 2014*, volume 8787 of *Lecture Notes in Computer Science*, pages 109–124. Springer International Publishing.

[Fader and Schmittlein, 1993] Fader, P. S. and Schmittlein, D. C. (1993). Excess behavioral loyalty for high-share brands: Deviations from the dirichlet model for repeat purchasing. *Journal of Marketing Research*, 30(4):pp. 478–493.

[Ghiassi et al., 2013] Ghiassi, M., Skinner, J., and Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications*, 40(16):6266 – 6282.

[Gundecha and Liu, 2012] Gundecha, P. and Liu, H. (2012). *Mining Social Media: A Brief Introduction*, chapter 2, pages 1–17.

[Hanzaee and Rouhani, 2013] Hanzaee, K. H. and Rouhani, F. R. (2013). Investigation of the effects of luxury brand perception and brand preference on purchase intention of luxury products. *African Journal of Business Management*, 7(18):1778–1790.

[Hardie et al., 1993] Hardie, B. G., Johnson, E. J., and Fader, P. S. (1993). Modeling loss aversion and reference dependence effects on brand choice. *Marketing science*, 12(4):378–394.

[Hu et al., 2014] Hu, P., Huang, M.-L., and Zhu, X.-Y. (2014). Exploring the interactions of storylines from informative news events. *Journal of Computer Science and Technology*, 29(3):502–518.

[Kervyn et al., 2012] Kervyn, N., Fiske, S. T., and Malone, C. (2012). Research dialogue. *Journal of Consumer Psychology*, 22(2):166–176.

[Kondreddi et al., 2014] Kondreddi, S., Triantafillou, P., and Weikum, G. (2014). Combining information extraction and human computing for crowdsourced knowledge acquisition. In *Data Engineering (ICDE), 2014 IEEE 30th International Conference on*, pages 988–999.

[Lindstrom, 2008] Lindstrom, M. (2008). *Brand sense: sensory secrets behind the stuff we buy*. Simon and Schuster.

[Manning et al., 2014] Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., and McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60.

[MLLib and SVM, 2015] MLLib and SVM (2015). https://spark.apache.org/docs/latest/mllib-linear-methods.html#linear-support-vector-machines-svms.

[Rafique, 2012] Rafique, M. (2012). Impact of celebrity advertisement on customers brand perception and purchase intention. *Asian Journal of Business and Management Sciences*, 1(11):53–67.

[Santos et al., 2016] Santos, R. F. D., Shah, S., Boedihardjo, A., Chen, F., Lu, C.-T., Butler, P., and Ramakrishnan, N. (2016). A framework for intelligence analysis using spatio-temporal storytelling. *GeoInformatica*, 20:285–326.

[Shukla et al., 2015] Shukla, M., Santos, R. D., Chen, F., and Lu, C.-T. (2015). Discrn: A distributed storytelling framework for intelligence analysis. *Virginia Tech Computer Science Technical Report http://hdl.handle.net/10919/53944*.

[Tang et al., 2015] Tang, S., Wu, F., Li, S., Lu, W., Zhang, Z., and Zhuang, Y. (2015). Sketch the storyline with charcoal: a non-parametric approach. In *Proceedings of the 24th International Conference on Artificial Intelligence*, pages 3841–3848. AAAI Press.

[van Rompay and Pruyn, 2011] van Rompay, T. J. L. and Pruyn, A. T. H. (2011). When visual product features speak the same language: Effects of shape-typeface congruence on brand perception and price expectations*. *Journal of Product Innovation Management*, 28(4):599–610.

[Wiersma, 2013] Wiersma, W. (2013). The validity of surveys: Online and offline.

[Zaharia et al., 2012] Zaharia, M., Chowdhury, M., Das, T., Dave, A., Ma, J., McCauly, M., Franklin, M. J., Shenker, S., and Stoica, I. (2012). 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)*, pages 15–28, San Jose, CA. USENIX.