

Visual Sentiment Analysis of Customer Feedback Streams Using Geo-Temporal Term Associations

ABSTRACT

Large manufacturing companies frequently receive thousands of web surveys every day. People share their thoughts regarding a wide range of products, their features, and the service they received. In addition, more than 190 million tweets (small text Web posts) are generated daily. Both survey feedback and tweets are underutilized as a source for understanding customer sentiments. To explore high-volume customer feedback streams, in this paper we introduce four time series visual analysis techniques: (1) feature-based sentiment analysis that extracts, measures, and maps customer feedback; (2) a novel way of determining term associations that identify attributes, verbs, and adjectives frequently occurring together; (3) a self-organizing term association map and a pixel cell-based sentiment calendar to identify co-occurring and influential opinion; and (4) a new geo-based term association technique providing a key term geo map to enable the user to inspect the statistical significance and the sentiment distribution of individual key terms. We have used and evaluated these techniques and combined them into a well-fitted solution for an effective analysis of large customer feedback streams such as web surveys (from product buyers) and Twitter (e.g., from Kung-Fu Panda movie reviewers).

Keywords: Customer Sentiment Visual Analytics, Term Association, Geo-Term Association, Pixel Geo Map, Key Term Geo Map, Pixel Calendar.

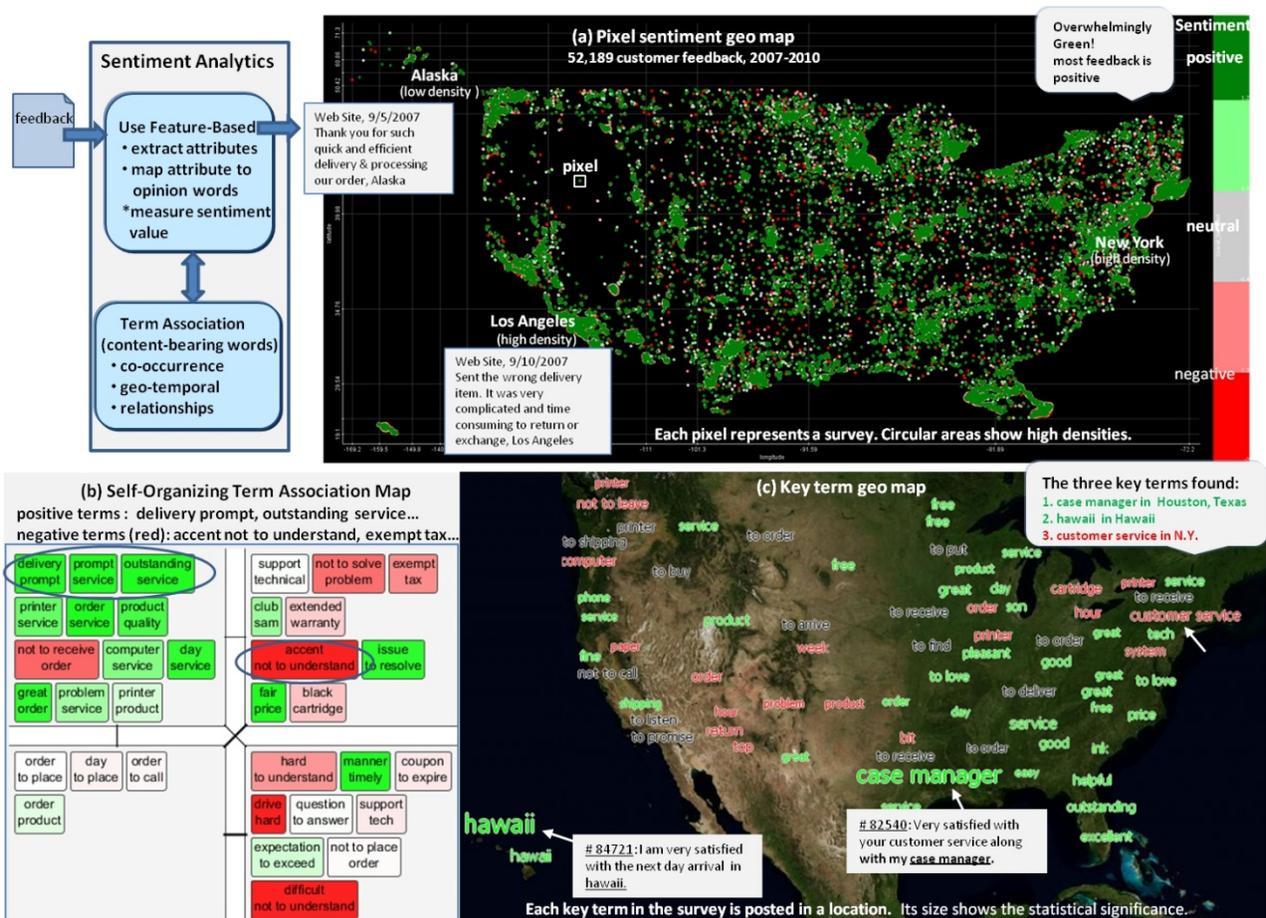


Figure 1: A pipeline using feature-based sentiment analytics and term associations with (a) pixel sentiment geo map, (b) self-organizing term association map, and (c) key term geo map for effective visualization of the sentiment distribution of the key terms.

1. Introduction

1.1 Motivation

With the rapid growth of social media, the number of customer comments available to corporations, business owners, and service managers interested in obtaining customer feedback is larger than ever. In addition to the traditional web survey, Twitter is a relatively new phenomenon that has the potential to generate massive amounts of customer comments. However, the language of the tweets is more casual than that of web reviews. Tweets are by definition short (maximum 140 characters) and tend to contain a significant number of abbreviations. The enormous size of the customer feedback data stream, the diversity of the comments, and the uneven distribution of feedback over time make sentiment analysis of this data very challenging.

A set of common questions arises in the analysis of customer comments from surveys and other online data streams. Are there aspects of location or geography that impact how a product or service is received by customers? Does a product or service work better for people on the coasts compared to people living in the interior of the country? Does it make a difference if the customer lives in a remote area rather than in a high density urban setting? Is the product or service more appreciated in certain states or cities? What are the important features, attributes, and associated context terms, such as products, timely delivery, channel vendors, product quality, or past experiences that our customers want? How significant terms (content-bearing words, e.g., compound nouns, adjectives, and verbs) are best extracted and presented to business managers so they can understand the results (positive versus negative)? Business managers want to see the sentiment value of the review, but they also want to know the important terms in the context of a review. Furthermore, how to visualize reviews in a dense area without overlap (e.g., Los Angeles, New York), is also a challenge required to be resolved.

To meet the above challenges, we propose a pipeline combining feature-based sentiment analysis and geo-term associations to enable store managers to analyze web survey feedback. As an example, Figure 1 shows customer feedback data from 2007-2010 in the United States as:

- a. Figure 1(a) is a pixel-based sentiment geo map that shows the sentiment results from 52,189 survey review. Each data point is a pixel representing a review. Color represents a sentiment value (green: positive, gray: neutral and red: negative). Figure 1(a) has overwhelmingly positive (green) reviews.
- b. Figure 1(b) is a self-organizing term association map used to visualize the relationships among each cluster of terms. E.g., prompt delivery, prompt service, and outstanding service frequently occur together as shown in the top left box of Figure 1(b).
- c. Figure 1(c) is a key term geo map used to display the most significant term in each geographical location. Color shows the average sentiment value of all the sentences containing the key term: e.g., “case manager” is the key term found in Houston, Texas. Its sentiment value is positive (green). New York has a key term “customer service” in red which results from people being concerned about their ability to understand the accent of the customer service representative.

1.2 Related Work

Much related work exists on analyzing twitter feeds. Bifet and Frank [12] proposed sliding window Kappa statistics for evaluation in time-changing data streams. Using these statistics, they performed a study on twitter data using machine learning algorithms for analyzing tweets. Marcus et al. [13] built a system, called TwitInfo, to perform automatic peak detection and labeling. TwitInfo allows users to browse a large collection of tweets using a timeline-based display that highlights the peaks of high tweet activity.

Feature-Based Sentiment Analysis

Feature-based sentiment analysis typically contains three successive steps: First, identify the attributes (features, i.e., nouns, compound nouns) customers commented on. Second, identify the sentiment words (i.e., good, bad), and third, map sentiment words to the attributes to which they refer. There are many different methods for extracting attributes: some use the frequency of terms that occur together in a sentence [7]; others use a certain threshold, e. g., Popescu et al. [6] consider all noun phrases as attributes whose frequency is above a certain threshold. To map an opinion word to an attribute, some of the methods [8] use distance-based heuristics, such as the closer a sentiment word is to the attribute, the higher its sentiment influence is on the attribute, or discrimination-based methods with a predefined word window [1]. Other approaches use natural language processing methods, such as Ng et al. [9] who use subject-verb and adjective-noun relations. We use a predefined set of syntactic reference patterns that are based on part-of-speech sequences [20]. In cases where this method is not able to resolve references, we use distance-based heuristics. For further analysis, we provide novel term association techniques to find the terms that frequently occur together (related work is described in Section 2.2).

Visual Feature-Based Sentiment Analysis

The most popular visualization for feature-based sentiment analysis is the tag-cloud [14] that visualizes reviews on the web. Maniwordle [15] provides users with flexible control over word clouds. Users are allowed to directly manipulate typography, color, position, and orientation for the individual words (e.g., attributes) as needed. SparkClouds [16] integrates sparklines [17] into a tag

cloud to visualize trends across a series of tag clouds. It simplifies line charts and gives users an overview of trends over time. Wanner et al. [18] visualize the development of RSS feeds over time that report on the U.S. elections. OpinionSeer [19] provides an interactive visualization system to analyze hotel customer feedback on the web using well-established scatter plots and radial visualizations. It displays opinion data inside a triangle. The radial visualization which is the bounding wheel of the opinion triangle is used for other data dimensions (i.e., time and location). The usefulness of the OpinionSeer depends on the volume of the reviews. For a large data volume it is hard to scale up, even with distortion, given the limited space inside the triangle.

In contrast to these approaches, we use feature-based sentiment analysis [1] combined with multi-resolution high density techniques [11] to process large customer feedback streams. We then analyze each feature to see if it is mentioned positively or negatively. Furthermore, we calculate term associations and construct a key term geo map to enable the analysts to quickly identify location specific differences between terms.

1.3 Our Goals and Contributions

In this paper, we present our approach to combining sentiment analysis [1] with a new term association technique as well as a geo-temporal visualization for an effective analysis of large customer feedback streams. To achieve this goal, **our first contribution** is that we introduce a feature and geo-based stream analysis technique that automatically detects which attributes (features) are frequently commented on, which attributes have interesting sentiment patterns, which attributes cluster significantly in certain geo-locations, and what terms (attributes, adjectives, verbs) often occur together. We then analyze each attribute to see whether it points us to interesting issues that customers have in general, or at certain points in time, or at certain geo-locations. In contrast to previous approaches, we identify term associations which consist of sets of content-bearing words, such as nouns, compound nouns, adjectives, and verbs, and are identified based on a sentence-wise co-occurrence. In addition, we evaluate which content-bearing terms are significantly associated with certain geo-locations.

Our second contribution is that we also propose two new geo-temporal visualizations (pixel sentiment and key term geo maps) that help users analyze large volumes of web surveys and twitter data. The sentiment pixel geo maps provide location patterns colored by the sentiment values from each feedback (red: negative, gray: neutral, green: positive). The locations with a large number of comments can be easily identified based on our circular pixel placement around the high density area, as shown in Figure 1(a) and Figure 12(b). We also developed a technique to visualize term associations using self-organizing maps (SOM) [21] as shown in Figure 1(b). SOM allows analysts to quickly identify which terms often co-occur; and related combinations of terms are clustered in one cell of the SOM. Key terms having biased geo-spatial distributions are automatically labeled on the map. In Figure 1(c), each label represents the most significant term discovered in a location, e.g., the term “case manager” in Houston, Texas. The customers in Texas like their “case manager” for solving their printer problems. In contrast, the feedback in New York shows that customers do not like the “customer service” due to language issues (difficult to understand). Furthermore, **our third contribution** is that a new pixel cell-based calendar is used by analysts to quickly discover temporal patterns based on time (e.g., hourly, daily, or monthly) as described in Section 3.2. The sentiment calendar is scalable both with respect to the number of comments and the number of attributes. We have applied these techniques to visualize web survey feedback (52,189 reviews), as shown in Figure 1, and tweets that are related to the movie Kung-Fu Panda (59,614 responses), and which are described in Section 4.2.

This paper is structured as follows: In Section 2, we describe the feature-based sentiment analysis and new term association techniques. An evaluation of association measures and n-ary associations and their strengths and weaknesses are also given in Section 2. In Section 3, we derive a suite of advanced visual analytics techniques: pixel sentiment geo map for visualizing dense areas without overlapping; self-organizing term association maps that show how terms are related; key term geo-maps for identifying the most significant terms for locations and the geo-distributions of term usage; pixel cell-based sentiment calendars for visualizing customer feedback and patterns over time. To validate the effectiveness of our techniques, section 4 presents two use cases; one with web survey data (from July, 2007 to June, 2011) and one with movie Twitter data. Section 5 concludes the paper and outlines our future work.

2 Our Approach

2.1 Feature-based sentiment analysis

In the literature the expressions *sentiment analysis* and *opinion analysis* are often used as synonyms. A sentiment or opinion is a statement that evokes either positive or negative associations. Often the attribute or feature to which a sentiment refers is of special interest. However, when analyzing open-ended data sources such as text streams, it cannot be accurately predicted which terms (features) will show up and therefore it is undesirable to use a predefined list of features. The analysis should be designed to be broad and cover all possibly interesting features. To this end we consider each noun or compound noun as a potential feature. Then, we use a *feature-based sentiment* algorithm [1] to

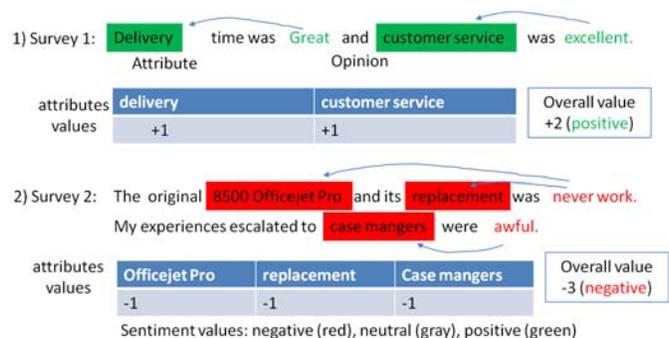


Figure 2: Methods to extract attributes and to measure sentiment values

measure the sentiment value, as shown in Figure 2. In addition, we store all other content-bearing words, such as verbs and adjectives, for further processing steps.

At times, certain features may have different sentiments in different contexts. This includes both the semantic context and the geo-spatial context of a feature within a review. In different contexts the sentiment that people associate with the same feature may vary. In addition, people in different regions may have different sentiments about the same topic. In order to account for these effects, we assess term associations within our data and visualize the geo-spatial distribution of features. More details are provided in the following sections.

2.2 Term Associations

The important terms of our document collections (features) have to be brought into context. To this end, information about associations between terms have to be automatically extracted from the text resources and visually conveyed to the analysts to enable better understanding of the data. The association strength of two terms can be measured regarding their sentence-wise co-occurrence. From an analytic point of view this task is closely related to frequent item set mining. However, the typical support and confidence approach is not very useful in the case of natural language, because term frequencies in text follow a long tail distribution as covered by Zipf's Law [4]. Some words are orders of magnitude more frequent and thus would be contained in many associations. Yet, highly frequent words usually carry less meaning than those with a moderate frequency and are thus not very valuable to explore. Brin et al. [3] consequently suggest relying on statistical measures for cases such as text data. Manning & Schütze [5] apply different statistical association measures to assess term co-locations: the hypothesis tests T-Test and Likelihood ratio as well as Pointwise Mutual Information (PMI). For the sake of brevity we refer interested readers to the referenced book for details about these methods. The assumption behind the hypothesis tests is the null hypothesis that two items are independent. If this hypothesis can be rejected with a high level of confidence the items can be considered to be associated. The more data points that support the rejection of the null hypothesis, the higher is the level of confidence.

To apply such methods for term associations we first have to define the probabilities that we work with. The Probability $P(a)$ that a term a occurs in a sentence s of the corpus C is defined as:

$$P(a) = \frac{|\{s : s \in C \wedge a \in s\}|}{|\{s : s \in C\}|}$$

The probability $P(a,b)$ that both term a and term b occur jointly in a sentence s of the corpus C is defined as:

$$P(a,b) = \frac{|\{s : s \in C \wedge a \in s \wedge b \in s\}|}{|\{s : s \in C\}|}$$

The above mentioned methods are applied to find the top binary associations, i.e., pairs of terms that are highly associated on a sentence basis. The performance of the different methods will be discussed in the evaluation section. For our analyses we included only terms that we consider being content-bearing, namely nouns, compound nouns, adjectives, and verbs. As mentioned earlier, the goal of extracting associations is to present them to the user with the intent of providing a more detailed insight into the results of the sentiment analysis. When extracting the top binary associations sometimes groups of associations show up that apparently belong together. For example, the top 100 associations from the web surveys contained {website, easy}, {website, to navigate}, and {easy, to navigate}. Evidently, these associations belong to the same frequently repeated statement "*website easy to navigate*" and should be merged. To this end we perform a form of apriori merging of binary associations to triples and then iteratively to sets of more than 3 terms until no further merging are possible. We found that the PMI is the only measure that can be extended in a straight forward manner to measure the association among more than two terms at a time. We calculate the PMI for $n \geq 2$ terms as:

$$I(a, b, \dots, n) = \log_2 \left(\frac{p(a, b, \dots, n)}{p(a)p(b) \dots p(n)} \right)$$

The prerequisite for getting an association containing a set of n terms is that all n distinct subsets containing $n-1$ terms are also considered to be associations. To give an example, an association {a,b,c} may exist if and only if {a,b}, {a,c} and {b,c} are considered to be associations. In addition, the following two requirements have to be fulfilled:

1. $I(a, b, c) > \max(I(a, b), I(a, c), I(b, c))$
2. $count(a, b, c) > lowerbound$

Where $count(a,b,c)$ denotes the number of sentences in the corpus that have to contain the three items jointly. This number has to lie above a certain user-defined threshold we name *lowerbound*. This threshold is necessary to prevent getting associations that are underrepresented. We denote this merging step as *PMI merging*. At times, the use of synonyms prevents sets from getting merged. For example, in the web survey dataset we get the associations {website, easy, to navigate} and {website, easy, to use}. Basically, both associations address the same statement just with slightly alternating expressions; some people say it is easy to use the website and some say it is easy to navigate. To cope with such usage of synonyms, associations containing more than 3 terms and sharing at least

50% of their terms are merged as well. We denote this step as *overlap merging*. To see whether both kinds of merging strategies for associations are beneficial to the analysis, we tested them for our data in the evaluation section.

After generating the associations, a sentiment value for each association is calculated. The process is slightly different for associations generated with *PMI merging* in comparison to associations generated with *overlap merging*. For an association generated with *PMI merging* a considerable number of sentences in the corpus exists (*>lowerbound*) that contain all terms of the association. For each of these sentences we sum up the sentiment values of all sentiment words of the association contained in the sentence. A positive word contributes +1 and a negative word contributes -1 to the sum. The average sentiment value of all sentences is considered to be the sentiment of the association. For associations generated with *overlap merging*, there might not exist a single sentence which containing all terms. Such an association is the composition of n overlapping associations generated with *PMI merging*. All sentences that contain at least one of the n overlapping associations are taken into account. The average sentiment value of these sentences is considered to be the sentiment of the association.

2.3 Geo Based Term Association

Mining term associations, as described in the previous section, enables the analyst to explore the semantic context in which a term has been used by the customers. In addition to the semantic context, terms can also be explored in their geo-spatial context. There is a whole set of geo-related analysis questions an analyst might have, such as "Do only customers in a certain location have a certain problem?" In order to shed light on previously hidden geo-distributional patterns in the customer feedback, we propose to mine term-location associations; that is, for each combination of a location and a term we apply the methods for hypothesis testing described in Section 2.2. This time, the null hypothesis is that the term is independent of the location. However, if a term can be observed more frequently within the feedback from a certain region than expected under independence assumption, the null hypothesis may be rejected with high statistical significance. In the latter case, the statistically most salient term-location associations can be conveyed visually for a closer inspection.

In the term-location analysis scenario the probabilities that we work with are different than the ones in the term association analysis described in section 2.2. First, we have the probability $P(x)$ that a term x occurs in a document d of the corpus C :

$$P(x) = \frac{|\{d : d \in C \wedge x \in d\}|}{|\{d : d \in C\}|}$$

Next, we have the probability $P(y)$ that a location y was the origin of a document d of the corpus C :

$$P(y) = \frac{|\{d : d \in C \wedge (\text{location}(d) = y)\}|}{|\{d : d \in C\}|}$$

Finally, the joint probability $P(x,y)$ that a term x occurs in a document of region y in the given corpus is defined as:

$$P(x,y) = \frac{|\{d : d \in C \wedge x \in d \wedge (\text{location}(d) = y)\}|}{|\{d : d \in C\}|}$$

2.4 Term Association Evaluation

Evaluation of Association Measures

It was not clear which of the outlined term association methods would perform best on real world data. Consequently, we applied and evaluated them. In addition to the T-Test, Likelihood Ratio Test, and PMI, we also applied a correlation coefficient (Phi). In order to get meaningful results we tested the methods on real data from web surveys. The data set consists of 52,189 responses to a customer web survey containing 96,987 sentences; the results are shown in Table 1.

T-Test	Likelihood Ratio	Phi	PMI
free, shipping (1741)	free, shipping (1741)	mouth, taste (8)	mouth, taste (8)
great, service (1929)	day, next (995)	not friendly, not to user (21)	74xl, 75xl (7)
excellent, service (1225)	order, to place (761)	club, sam (21)	bang, buck (6)
day, next (995)	great, service (1929)	expectation, to exceed (51)	office home, student (6)
order, to place (761)	excellent, service (1225)	creative, kit (12)	god, to bless (8)
to keep, work (480)	to keep, work (480)	74xl, 75xl (7)	aol, yahoo (6)
good, work (494)	day delivery, next (313)	manner, timely (82)	creative, kit (12)
fast, service (599)	hour, phone (416)	free, shipping (1741)	not friendly, not to user (21)
free, next (523)	good, work (494)	bang, buck (6)	bait, switch (6)
hour, phone (416)	hour, to spend (268)	office home, student (6)	citizen, senior (10)

Table 1 The top 10 binary associations for the web surveys generated with each measure. Only pairs of terms co-occurring in at least 6 sentences were considered. The number of co-occurrences of each term pair is put into parentheses.

The results in Table 1 show that the two hypotheses tests tend to prefer rather frequent associations, whereas the two other measures tend to find more infrequent associations that are less general. In order to gain further insight we examined the frequency distribution among the top 100 associations. Figure 3 shows the distribution for the web surveys.

PMI and Phi prefer rather infrequent associations. Therefore, we regard both measures as not very suitable for our task. The T-Test, in contrast, especially for the large data set, tends to prefer associations with a very high frequency. The Likelihood Ratio Test is the only measure that covers almost the whole frequency spectrum. In a more detailed analysis, we found that the Likelihood Ratio Test is the best choice for our approach, as highly frequent associations are more interesting in the general case, although there are still many rather infrequent associations that lead to interesting findings. For mining geo-term associations, accordingly, we also preferred the Likelihood Ratio Test.

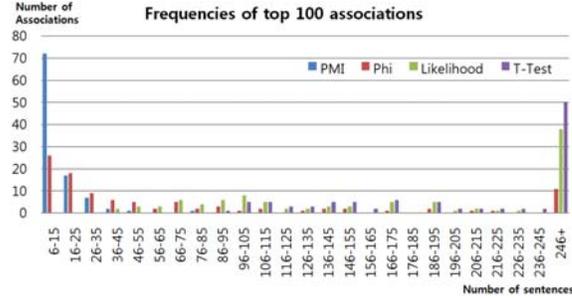


Figure 3: The frequency distribution of the top 100 associations extracted with each measure for the web surveys. The y-axis shows the frequency of associations in the corpus, i.e., in how many sentences associations occur. The x-axis reveals how many of the top 100 associations had a certain frequency.

Evaluation of n-ary Associations

To evaluate the performance of the suggested merging steps, we applied them to our data. Additional merges were achieved through overlap merging. The results are shown in Table 2.

Web Surveys: Top 10 associations after PMI merging	Web Surveys: Top 10 associations after PMI merging and overlap merging
door, front, to leave (27)	free, overnight, shipping, price, delivery, fast, to love, appreciate
good, to keep up, work (55)	english, someone, to speak, people
hour, phone, to spend (154)	good, to keep up, work, to keep
english, someone, to speak (30)	easy, to navigate, website, to use
address, to deliver, wrong (63)	day shipping, free, next, day delivery
english, people, to speak (39)	address, to deliver, wrong, fedex
easy, to navigate, website (57)	day, next, to receive, to order
good, to keep, work (385)	hour, phone, to spend
courteous, helpful, knowledgeable (9)	door, front, to leave
day ship, free, next (153)	fedex, package, to leave

Table 2 The merging results for the two datasets. The absolute number of sentences containing an association are in parentheses (applies only to the left side).

The n-ary associations are very useful. Often they can readily be interpreted as a statement, e.g. {easy, website, to use} indicates that the website is easy to use. Also, the overlap merging produces nice results. For example, {good, to keep up, work} and {good, to keep, work} were merged into one association {good, to keep up, work, to keep}. In some cases our preprocessing algorithms were just not able to find the particle “up” and relate it to “keep”. This problem is now partly solved by merging terms together in the term association step.

3. Visual Analytics

3.1 Geo Maps

Pixel Sentiment Geo Map

Sentiment analysis of customer feedback is a process that often excludes geo-spatial information. The sentiment analysis process mainly focuses on how the customers like or dislike an object and what attributes of the object the customers commented on. Only a few analyses focus on the spatial distribution of opinions and show the influence of the geographic locations towards the sentiment. However, it is desirable to take geographic location into account as it may influence the sentiment in customer feedback. In a marketing process, for example, it may be important to analyze why the people of a particular area did not like a movie or a product. Geographically aware sentiment analysis may enable new insights into the reasons for success or failure of a service or a product and lead to design variants of a product that are customized to local preferences.

Adding the geographic information of opinions to the analytical process makes things more complicated. As soon as we deal with the locations of user-generated data, we encounter different data densities resulting from varying population distributions. The unequal distribution of data complicates the display of data. Overlap often causes the loss of important information, such as the distribution of opinions within a region. A frequent approach is to cluster the data spatially and show the aggregation of the underlying data for each cluster, for instance, the average sentiment or the distribution of opinions by graphics or small bar charts. A severe drawback of this method is the disappearance of the original data points and the creation of visual artifacts due to the binning and aggregation processes. The insights gained from these visual representations may be biased by incorrect clustering or the aggregation method used.

We propose another way to visualize all the data points seen which avoids overlap. We apply a pixel placement algorithm to the data to avoid overlapping data points (reviews). Our pixel placement algorithm replaces the overlapping points by a circle of points positioning them at the nearest free position within the circle. Figure 4 shows the algorithm which is based on the method presented in [2]. The result of this technique is a visualization that shows each single data point as shown in Figure 1(a). The pixel sentiment geo map shows the sentiment distribution of recent buyers responding to a web survey. The color of each pixel (review) represents its sentiment value (red: negative (≤ 0); gray: neutral (0); green: positive (≥ 0)). A high density area, such as in Los Angeles and New York, is identified by a circle with non-overlapping reviews placed around it. Each review in the geo-map is accessible; users can mouse over a review and read the content, such as the term association, e.g. {next day, arrival} appearing in the positive feedback from Hawaii as shown in Figure 1(c).

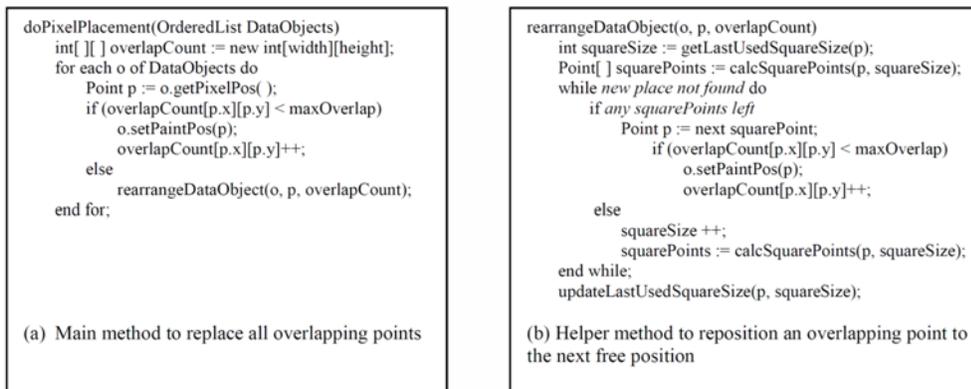


Figure 4: The algorithm used for pixel placement to ensure a visual representation of points without overlapping

Our algorithm displaces the points in the order of their priority (e.g., the sentiment of the point) to avoid random patterns in the resulting visualization. In order to avoid overlapping we have to remember which pixel locations are already occupied; therefore we need a two-dimensional integer array representing each pixel of the display area. For each data point, the program has to look up the number of data objects already placed at the preferred position of the data object and compare this to the maximum allowable number of overlapping points; in our case we set this value to 1 as we allow one data point per pixel maximum. If the current data object can be placed at its original location, we store this information in the two-dimensional integer array. Otherwise, we have to look for the nearest free pixel position in order to place the current data object there, as illustrated in Figure 4. The procedure `rearrangeDataObject` does the real pixel placement: In order to speed up our algorithm we store the radius for each pixel that was used for the last displacement (The initial value is 1). We can calculate the pixels of the circle around point p with this radius. The determination of the next free pixel location is done based on a modified version of the Bresenham-Midpoint [10] algorithm using a line width of two.

The pixel placement approach is sketched in Figure 5 and looks at the placement of a data object in the pixel placement process. Just assume that the current data object originally is located at the pixel position marked with a black X. As this position is already occupied by some other previously processed data objects we circularly iterate around the original position until we find the next free position. The possible free positions are the ones marked with a green color and result from the Bresenham-Midpoint algorithm described above.

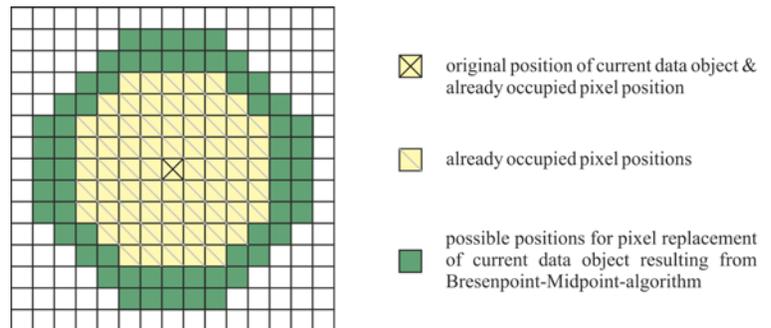


Figure 5: Schematic explanation of pixel placement process

Key Term Geo Maps

The geo-spatial information available for the web surveys is the zip code of the customer's address. The zip codes can be mapped to zip code areas on a map. In principle, it would be interesting to see which terms are associated with which zip code areas. However, in the web surveys we have to deal with a data sparseness issue: On average we have less than one review per zip code area. With the data being so sparsely scattered, reliable statistics cannot be derived. Consequently, we have to change the granularity of analysis. Figure 6 shows different levels of granularity both with respect to the text and the geo-spatial component. Term-location associations may be calculated for any combination of these. For the web survey dataset we determined "key term - county" associations to be useful. However, if the analysis is supposed to focus on more general aspects, it can also be evaluated which topics are associated on the state level.

We suggest two complementary visualizations to enable the visual analysis of term-location associations: (1) the Key Term Geo Map (see Figure 1c) and (2) the Key Term Distribution Map (see Figure 7). The data used to discuss both visualizations are from the customer web surveys.

(1) Key Term Geo Map: For each text unit (e.g., key term) and each geo-spatial unit (e.g., county) we calculate the association value according to the Likelihood Ratio Test as described in Section 2.4. The term-location associations with the highest significance values are displayed on a map. We iterate through the list of term-location associations ordered by descending significance values. In each step, we try to place the current term at its respective location without causing labels to overlap. If some overlap were introduced, we skip the current term-location association and proceed with the next one. Otherwise, we place the label on the map in the color of its sentiment value and size the label according to the significance value of the association. As we zoom into the map, there will be more labels (terms) shown as the available space increases.

(2) Key Term Distribution Map: Our second visualization enables the user to inspect the sentiment distribution of individual key terms. When a key term is selected by the user, a new view is created conveying information for this key term only. More specifically, we first determine all occurrences of the selected key term and retrieve the respective sentiment value. The data is then used to generate the key term distribution map as shown in Figure 7. We first partition the data into two subsets: the occurrences with positive sentiment in Figure 7(a) and occurrences with negative sentiment in Figure 7(c). The two partitions are processed separately. We apply a Gaussian blurring function in order to spatially extend the occurrences and increase the visual salience of distribution patterns. We thus obtain a blurred representation for both sentiments showing the respective occurrences of the selected term as depicted in Figures 7(b) and 7(d). Finally, we generate a combined image by using the RGB channels of the RGB color model. The blurred image of the negative occurrences is put in the red channel and the green channel is used for the positive occurrences. Consequently, locations with both positive and negative sentiments will result in yellow colors, while pure positive sentiments will result in green colors. The final result of our technique can be seen in Figure 7(e).

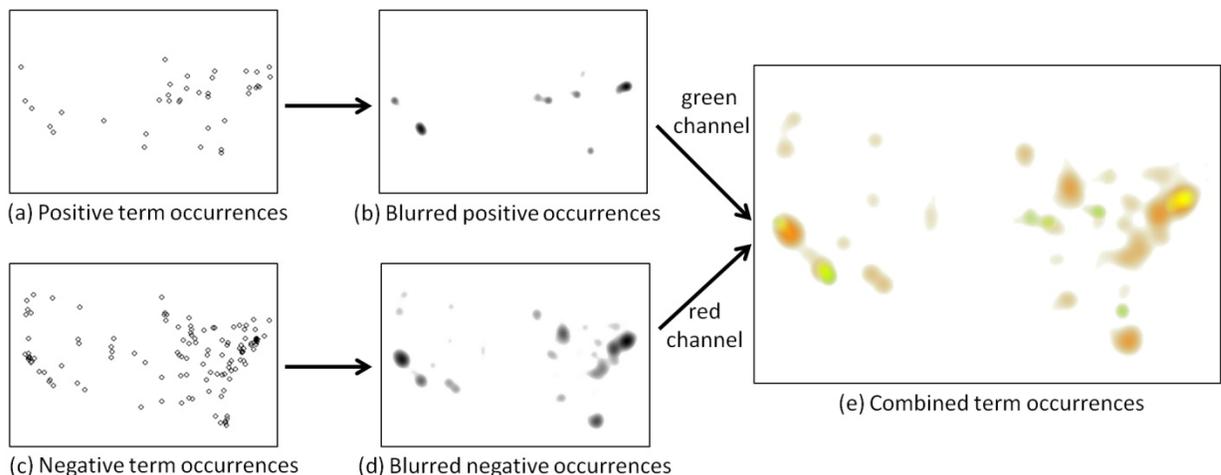


Figure 7: Generating the key term distribution map: We blur the locations of positive (a) and negative (c) term occurrences resulting in (b) and (d). Combining both results by using RGB channels will produce our key term distribution map, shown in Figure 7(e).

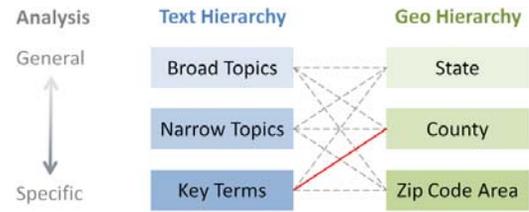


Figure 6: Term-location associations can be calculated using combinations of different granularities.

3.2 Pixel Cell-Based Sentiment Calendar

Figure 8 shows a monthly calendar view with sentiment attributes generated from the buyer's web survey data. This calendar is defined by an x-axis (day), a y-axis (year and month), and a color (sentiment value). Each pixel cell represents a review. Business managers can quickly observe the variances e. g.; printer and website have more red than delivery and shipping. Business managers can rubber-band the area on 11/2009, days 8, 9, and 10 and query on finding the geo locations of the negative comments (Figure 8(b)) and other attributes which have a high correlation with "printer" (Figure 8(c)). In the correlation window, service managers can easily mouse over a colored pixel to read the review content, e.g., on 11/9 at the first red pixel: "Printer support is great but this printer gobbles ink". This observation validates that the terms {printer, ink} often occurred together.

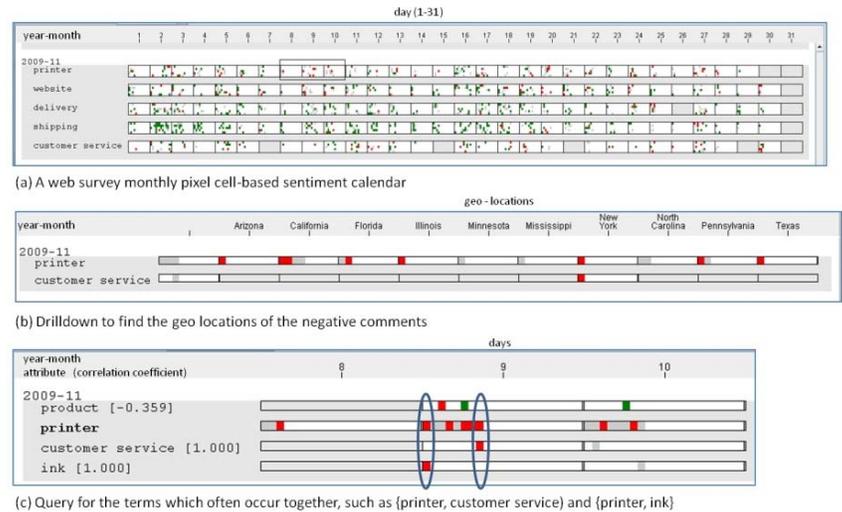


Figure 8: A monthly sentiment calendar

- (a) A web survey (product buyers) monthly pixel cell-based sentiment calendar
- (b) Users are able to rubber-band around 9/8/2009 and drilldown to find the geo locations of the negative "printer" feedback.
- (c) Users can issue a query to locate the terms associated with the attribute "printer".

3.3 Self-Organizing Term Association Map (SOM)

Section 2.2 describes how individual words are grouped into associations based on their sentence-wise co-occurrence. One association represents one problem; for example in [the web survey data stream collected from monthly historical data](#), the association {address, to deliver, wrong, fedex} summarizes the complaints of customers that FedEx delivered their order to the wrong address. In many cases, such an interpretation of associations is quite obvious. However, in some cases it is still valuable for the analyst to have quick access to the sentences or whole reviews that contain an association to understand or verify the meaning. Therefore, we provide information about the associations in an interactive visual interface. Instead of simply listing associations we want to enrich them with further information. As illustrated in Figure 1(b), we color each association with its sentiment value, i.e. the average sentiment of sentences containing the association. Positive sentiments are mapped to green and negative sentiments to red; the color saturation indicates the sentiment strength. Furthermore, we cluster associations according to the reviews to which they belong. While the associations can be interpreted as statements extracted from sentences, the association clusters can be interpreted as groups of statements often made within the same reviews. For the clustering a distance measure between two associations has to be defined. To do so, we create a high-dimensional vector for each association that has as many dimensions as there are reviews in the data set. If an association is contained in a specific review, the entry in the respective dimension will be 1; otherwise, it will be 0. To calculate the distance between two associations we take the Euclidean distance between their vectors.

Instead of computing separate clusters of associations, we want to reflect how the clusters relate. With respect to negative associations one cluster is dominant. This cluster on the top right deals with problems regarding the language skills of the customer support teams. Some customers find the accent difficult to understand. A dominant positive feedback is easily analyzed. In the top left cluster, people like the service and especially the prompt delivery.

In comparison to the standard "word cloud" visualization, the additional structure provided by the term associations gives more insights by enriching words with semantic context information. However, the SOM visualization also reveals some limitations of the overall approach. When the real number of clusters in the data is larger than the number of SOM nodes, some SOM nodes necessarily show a mixture of several topics. In addition, preprocessing errors may also be revealed. For example, when hovering over the association {hard, drive} it can be seen that people do not have "hard times with their drives" as the strongly negative sentiment would suggest. They are simply making a comment about their "hard drive", which is neither negative nor positive. The misleading representation is due to the fact that the preprocessing algorithm failed to detect "hard drive" as a compound noun and interpreted "hard" to be a sentiment referring to "drive".

4. Use Cases and Evaluations

The combination of sentiment analysis and term associations with the above visual analysis techniques has a large number of applications, including hotel reservations, product surveys, IT services, theme park attractions, movies, etc. To validate our approach,

we have used two data streams: web surveys and Twitter data. Web surveys are historical data which collected after customer purchased products. Twitter data is collected in real time through HTTP connection to the Twitter API. We use ingestion adapters pull data by specifying different keywords (e.g., kung Fu Panda, Hangover, etc.) from difference sources.

4.1 Web Survey Data Streams

Pixel Sentiment Geo Map

As illustrated in Figure 1, population density sometimes overshadows other aspects. For example, New York and Los Angeles have dense populations and hence are likely to produce many comments. Our solution allows all comments, even in high density areas to be visualized and explored. Figures 9 (A) and (B) illustrate the value of the geospatial analysis. Customers in Los Angeles are relatively unhappy with the delivery as compared to customers in a sparsely populated area such as Alaska. Extracting opinion associations for particular regions may provide insights into regional preferences and needs.

Key Term Geo Map

The key term geo map is used to identify which term is the most significant in a location. The font size reflects the significance value, and the color represents the sentiment value. Analysts may zoom into the map to visualize more terms, e.g., zooming from Figure 1(c) (USA) to Figure 10 (Houston or Hawaii). The strongest association, somewhat surprisingly, is the term “case manager” in Harris County, Texas, as shown in Figure 10(A). The second strongest association was: the term “hawaii” significantly associated with Hawaii County as shown in Figure 10(B). From the map, the analyst can quickly detect spatial patterns of term usage and explore the causes by mousing-over the term to see the full customer comments. Finally, the business manager may decide to improve the sales and service policies according to the uncovered causes.

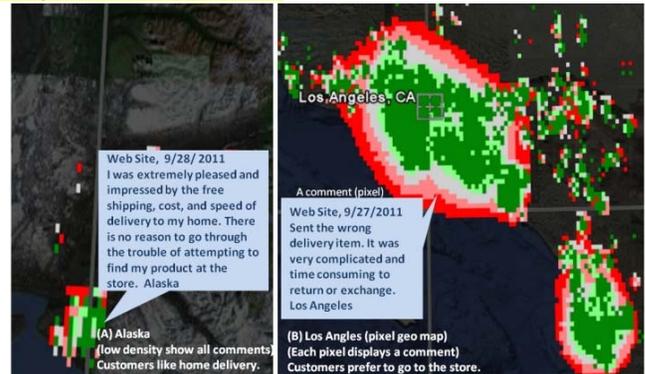


Figure 9: A comparison of comments from different areas (A) Comments in Alaska; and (B) Comments in LA.



(A) Term “Case Manager” at Harris County
82540: Very satisfied with your customer service along with my case manager.



(B) Term “hawaii” at Hawaii County
#39997: I am very satisfied with the next day arrival in hawaii. That’s great. Thanks.

Figure 10: Key Term Geo Map

Key Term Distribution Map

As described in Section 3.1 the analyst can select a key term and the geo-sentiment patterns for this selection will be displayed on a map. Figure 11 shows some interesting distribution patterns:

- print cartridge*: Different positive and negative local clusters appear indicating customer feedback on printer cartridge.
- shipping*: The sentiments for shipping are equally distributed over the whole map, which indicates that negative and positive comments are mostly balanced.
- delivery service*: Appreciation of delivery service, on the other hand, is not distributed the same across all areas.
- delay*: Delayed deliveries seem to be more likely in certain areas. Interestingly, the most negative area correlates with the area where people complain about traffic (see Figure 11(h)).
- sales tax*: The complaints about sales taxes show localized clusters.

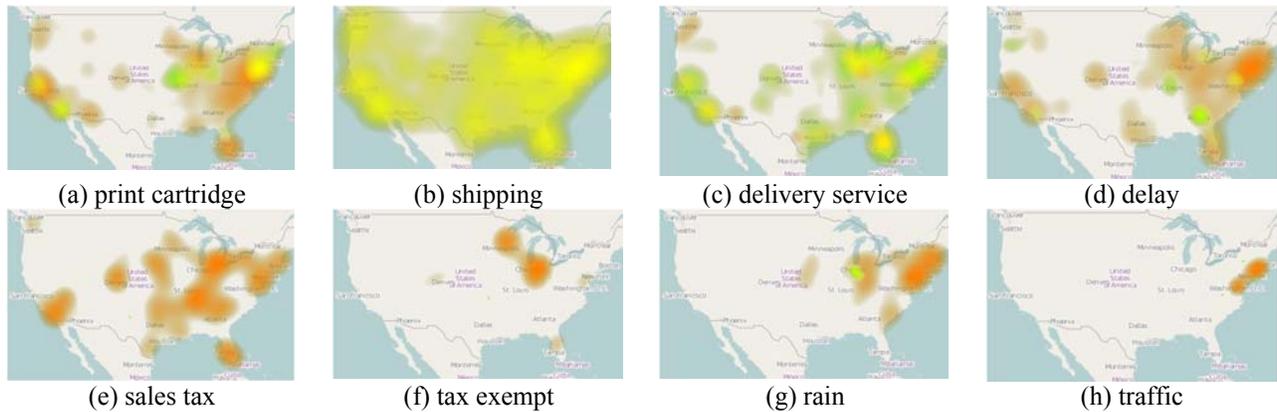


Figure 11: The geo-sentiment distribution for different attributes

- (f) *tax exempt*: Similarly, problems regarding tax exempt buyers also appear in certain areas only. Interestingly, this issue and the sales tax issue both also show a burst pattern over time. This indicates that the issue only occurred at certain locations at a certain point in time and has apparently been resolved.
- (g) *rain*: As expected weather phenomena are not the same over the whole map. Only in certain areas people complained that their packages were left in the rain.
- (h) *traffic*: Complaints about traffic is concentrated in certain East Cost areas. The complaints about traffic appear to be concentrated in areas similar to the areas associated with delays as shown in Figure 11(d).

4.2 Movie: Kung-Fu Panda Twitter Stream

Gain better spatial insights from geo sentiment map

We used our geo-sentiment map to analyze the sentiments towards the Kung-Fu Panda movie during the opening week. Each data point represents a person's comment about the movie and indicates a feature they liked or a feature they did not like. The map reveals several dense areas that indicate a large number of reviews posted on Twitter. Overall there were 59,614 tweets about Kung-Fu Panda from all the geographic locations available. There are a number of high-density areas each with a large number of tweets that resulted in highly overplotted regions. Using our pixel placement approach we are able to avoid the overlap. The sentiment pixel geo map allows us to visualize large numbers of data fitting entirely into the display window without any overlap.

To evaluate the effectiveness of this geo sentiment map, we compare it (Figure 12(b)) with the ordinary map shown in Figure 12(a). In Figure 12(a) we show a visual representation of the twitter data on a map with data-induced overlap. The problem is that the density and value distribution may vary in a region, which may not be visible due to overlapping pixels. The geo sentiment map in Figure 12(b) has no overlap with each single tweet being represented as one pixel by applying our pixel placement algorithm. Users are able to navigate through the dense areas for further analysis and see each tweet in detail along with the calculated sentiment. Further analysis of the sentiment distribution can lead to a better understanding of how this movie was received in various locations.

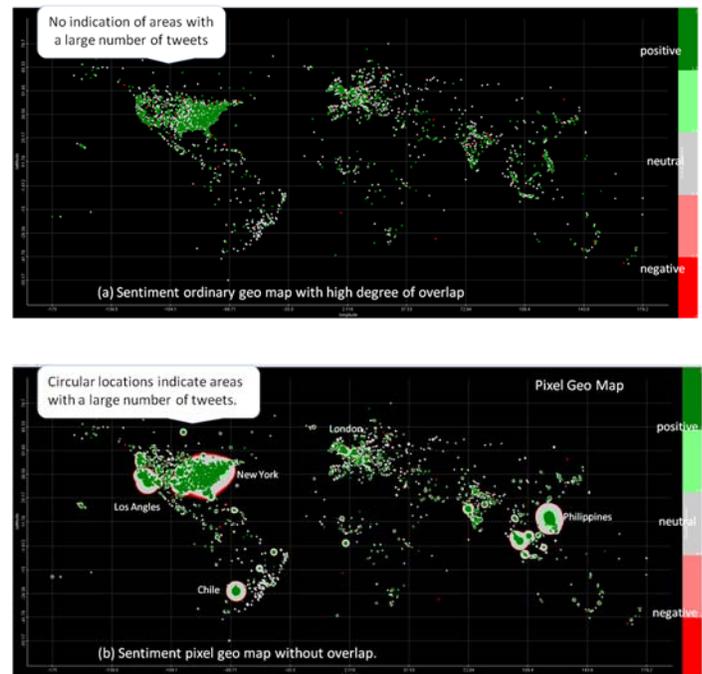


Figure 12: Geo map high density area evaluation (e.g., New York and Los Angeles)

5. Conclusion

With the currently available high speed and high volume customer feedback streams, new sentiment techniques are needed for helping companies learn what their customers like or dislike about their products and services in real-time. In this paper, we presented a novel integrated suite of methods that covers the whole analysis pipeline. First, we employ a feature-based algorithm to extract attributes, find opinions, and measure their sentiment values. Then, we extend the sentiment analysis to term associations. Our novel sentence-based term association algorithm and measurement methods can quickly identify the terms (i.e., nouns, verbs, and adjectives) that occur frequently together. Our combined analysis approach extends the scope of the customers' sentiment information that business managers should know about. In visualizing such a large volume of feedback data, there are three main issues: scalability, density, and context dependency. To solve these problems, we introduced pixel sentiment geo-maps and pixel sentiment calendars. Using a pixel sentiment geo-map, analysts can gain better insights from geographical sentiment distributions and are able to quickly identify areas of interest such as high density areas. Using the key term geo map, analysts can easily identify the most significant local feedback terms and their volume. Using the key term distribution map, analysts can visually explore different attributes for interesting geo-distributional patterns. Using a pixel sentiment calendar, analysts can gain better insights into temporal patterns of a large customer feedback stream. From our experiments, we see that in some cases population density may overshadow sentiment aspects. To visualize hundreds of terms in a single view, we introduce a variant of a self-organizing map which clusters related terms into related nodes. The color of a term association represents the aggregated sentiment value of all contained comments. From the sentiment value, analysts can quickly identify the important terms and initiate the proper actions.

The combined techniques mentioned above have been successfully employed in analyzing a number of use cases, including hotel reviews, movie tweets, and web surveys. We have discovered numerous customer concerns and initiated corresponding improvements. Our future work will proceed to detect geo-temporal sentiment patterns, trends, and influences in the customer feedback streams for live alerts.

ACKNOWLEDGMENTS

The authors wish to thank Malu Castellanos and Riddhiman Ghosh for providing the Kung-Fu Panda tweets and their comments and suggestions.

REFERENCES

- [1] D. Oelke, M. Hao, C. Rohrdantz, D. A. Keim, U. Dayal, L. E. Haug, H. Janetzko. Visual opinion analysis of customer feedback data. VAST09
- [2] Keim, D. A., Hao, M. C. Dayal, U., Janetzko, H., and Bak, P., Generalized Scatter Plots. *Information Visualization Journal (IVS)*, 2009.
- [3] Sergey Brin, Rajeev Motwani, and Craig Silverstein. 1997. Beyond market baskets: generalizing association rules to correlations. In *Proceedings of the 1997 ACM SIGMOD international conference on Management of data (SIGMOD '97)*, Joan M. Peckman, Sudha Ram, and Michael Franklin (Eds.). ACM, New York, NY, USA, 265-276.
- [4] Zipf, George K.: *Human Behaviour and the Principle of Least Effort*. Cambridge: Addison-Wesley Press (1949).
- [5] Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. The MIT Press, First edition, June 1999.
- [6] A.-M. Popescu and O. Etzioni. Extracting product features and opinions from reviews. In *HLT '05: Proc. of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 339–346. Assoc. for Computational Linguistics, 2005.
- [7] X. Ding, B. Liu, and P.S. Yu. A holistic lexicon-based approach to opinion mining. In *Proc. of the international Conference on Web Search and Web Data Mining (WSDM '08)*. ACM, pages 231-240, 2008.
- [8] Ding, X., Liu, B., and Yu, P. 2008. A holistic lexicon-based approach to opinion mining. In *Proceedings of the international conference on Web search and web data mining*. 231-240.
- [9] Ng, V., Dasgupta, S., and Arifin, S. M. N. 2006. Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews. In *Proceedings of COLING/ACL 2006 Main Conference Poster Sessions*. 611-618.
- [10] Bresenham, J. A linear algorithm for incremental digital display of circular arcs. *ACM Volume 20 issue (2)*, page 100–106. 1977.
- [11] Hao, M. Dayal, U. Keim, D. A., Schreck, T., Multi-Resolution Techniques for Visual Exploration of Large Time-Series Data. *Proceedings: IEEE VGTC Symposium on Visualization, EuroVis 2007*.
- [12] Bifet A., Eibe F. Sentiment Knowledge Discovery in Twitter Streaming Data. University of Waikato, Hamilton, New Zealand. 2011.
- [13] Marcus A., Berstein M. Badar O. et al. TwitInfo: Aggregating and Visualizing Microblogs for Event Exploration. Vancouver, Canada. 2011.
- [14] Viégas, F. B., Wattenberg, M. Feinberg J., Participatory Visualization with Wordle, *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, pp. 1137-1144, 2009.
- [15] Joy, k. lee B. et al. ManWordle: Providing Flexible Control over Wordle. *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, No. 6, 2010.
- [16] Lee, B, Riche, N., et al. SparkClouds: Visualizing Trends in Tag Clouds, *IEEE Transactions on Visualization and Computer Graphics*, vol. 16,
- [17] Tufte, E.R., *Beautiful Evidence*, Graphics Press.
- [18] Wanner, F. Rohrdantz, C., Mansmann, F., Oelke, D., and Keim, D. A. Visual sentiment analysis of RSS news feeds featuring the US presidential election in 2008. In *Workshop on Visual Interfaces to the Social and the Semantic Web (VISSW 2009)*, 2009.
- [19] Wu, Yingcai, Wei, Furu, Liu, Shixia, Au, Norman, Cui, Weiwei, Zhou, Hong, and Qu, Huamin. 2010. OpinionSeer: Interactive Visualization of Hotel Customer Feedback. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (November 2010).
- [20] S. Kisilevich, C. Rohrdantz and D. A. Keim. "Beautiful picture of an ugly place". Exploring photo collections using opinion and sentiment analysis of user comments. In *Computational Linguistics & Applications (CLA 10)*, 419–428, 2010.
- [21] Kohonen, Teuvo, Self-Organizing Map, *Proceedings of the IEEE*, vol. 78, No 9, September 1990.
- [22] Gorman, S.: The Oscars and Location Based Sentiment Analysis through Twitter. February, 2011.