

Integrating Sentiment Analysis and Term Associations with Geo-Temporal Visualizations on Customer Feedback Streams

Ming Hao¹, Christian Rohrdantz², Halldór Janetzko², Daniel Keim², Umeshwar Dayal¹
 Lars-Erik Haug¹, Mei-Chun Hsu¹
¹Hewlett Packard Labs, USA, ²University of Konstanz, Germany

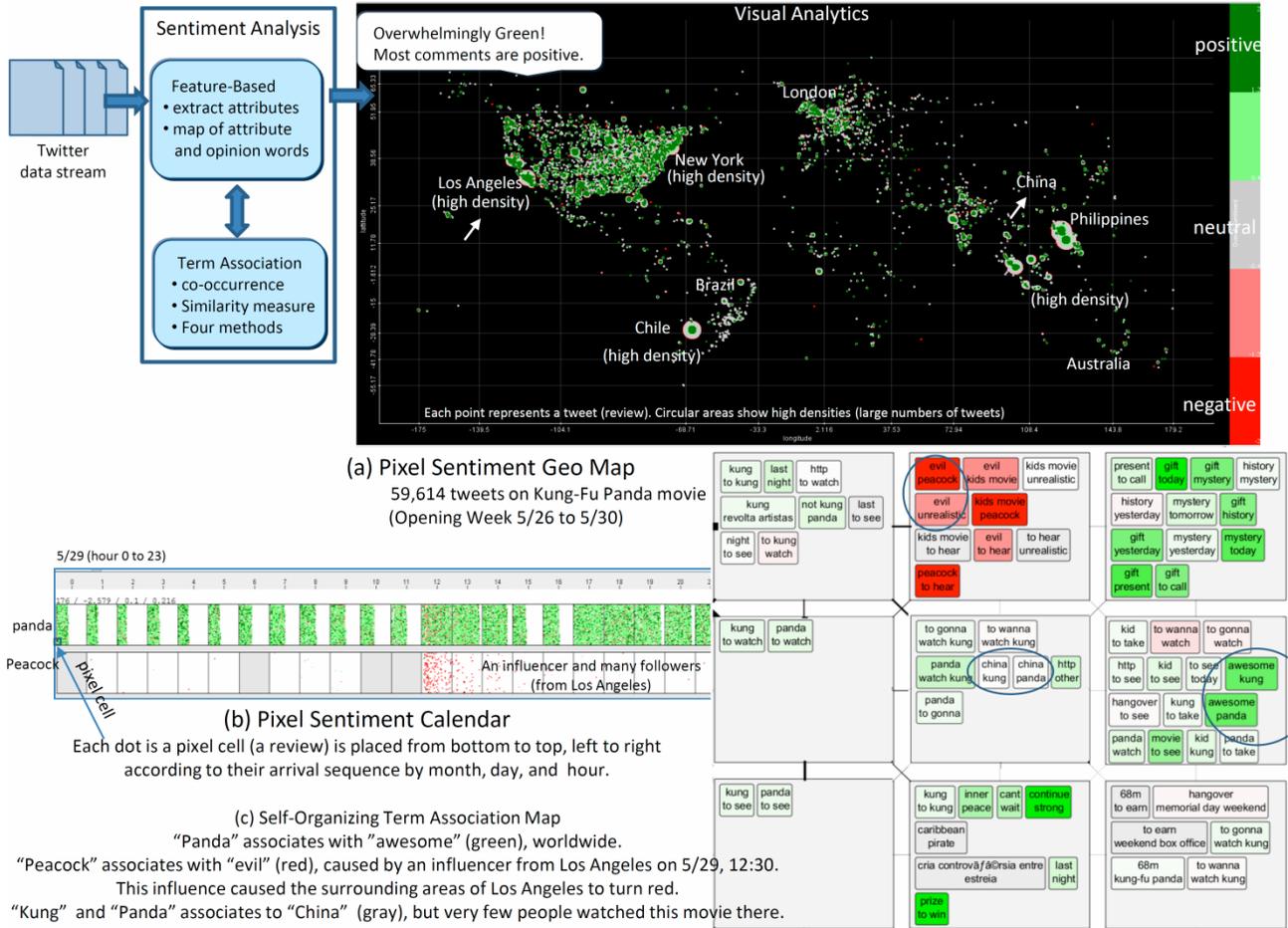


Figure 1: A pipeline of integration sentiment analysis and term associations with geo-temporal visualizations for effective visualization of large customer feedback streams from Twitter

ABSTRACT

Twitter currently receives over 190 million tweets (small text-based Web posts) and manufacturing companies receive over 10 thousand web product surveys a day, in which people share their thoughts regarding a wide range of products and their features. A large number of tweets and customer surveys include opinions about products and services. However, with Twitter being a relatively new phenomenon, these tweets are underutilized as a source for determining customer sentiments. To explore high-volume customer feedback streams, we integrate three time series-based visual analysis techniques: (1) feature-based sentiment analysis that extracts, measures, and maps customer feedback; (2) a novel idea of term associations that identify attributes, verbs, and adjectives frequently occurring together; and (3) new pixel cell-based sentiment calendars, geo-temporal map visualizations and self-organizing maps to identify co-occurring and influential opinions. We have combined these techniques into a well-fitted solution for an effective analysis of large customer feedback streams such as for movie reviews (e.g., Kung-Fu Panda) or web surveys (buyers).

Keywords: Sentiment Analysis, Term Associations, Geo-Temporal Visualization, Feedback Streams.

1. Introduction

1.1 Motivation

With the rapid growth of social media, the number of customer comments available to corporations, business owners, and IT service managers interested in obtaining customer feedback becomes larger than ever. Figure 1 shows Twitter comments on the movie Kung-Fu Panda. During the five days from 5/26 to 5/30, there were 59,614 tweets related to this movie. This number was much larger than that of traditional web reviews. The tweets came in by the minute in contrast to the normal web reviews that come in on a daily basis. However, the language of the tweets is more casual than that of web reviews, and twitter comments are much shorter (maximum 140 characters), containing a significant number of abbreviations. The enormous size of the data stream, the diversity of the comments, and the uneven distribution of tweets over time make the analysis of twitter data very challenging.

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 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. In contrast to these approaches, we use feature-based sentiment analysis [1] with multi-resolution high density techniques [11] to process large customer feedback stream in real-time. We then analyze each feature to see if it is mentioned positively or negatively.

Feature-Based Sentiment Analysis

Feature-based sentiment analysis often 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 to extract the attributes: some use the frequency of terms that occur together in a sentence [7]; and some 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 heuristic. 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* [18] visualizes 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 with the limited space inside the triangle, even with distortion.

1.3 Our Goals and Contributions

In this paper, we present our approach to integrate sentiment analysis and term associations with geo-temporal visualization for an effective analysis of large customer feedback streams. To achieve this goal, we develop a novel feature-based stream analysis technique that automatically detects which attributes (features) are frequently commented on, which attributes have similar negative comments, and which sources (locations) and what terms (attributes, adjectives, verbs) often occur together. We then analyze each attribute to see if it is mentioned positively or negatively. In contrast to previous approaches, we identify term association after each attribute has been properly scored. The term associations are content-bearing, consist of nouns, compound nouns, adjectives, and verbs, and are identified based on a sentence-wise co-occurrence. We have applied these techniques to visualize both Kung-Fu Panda movie tweets (59,614 reviews), as shown in Figure 1, and customer web survey data (47,966 responses) as described in subsection 4.2.

We also propose several geo-temporal visualization techniques that help the users to analyze large volumes of twitter data and web surveys. Our sentiment pixel geo-maps provide location patterns colored by the sentiment values on 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. To identify temporal patterns, in contrast to other visual summarization techniques, a pixel cell-based calendar is used for analysts to quickly discover temporal patterns based on time (e.g., hourly, daily, and monthly). The sentiment calendar is scalable both with respect to the number of comments and the number of attributes. In addition,

we also developed a technique to visualize term associations using a self-organizing map (SOM) [21]. These maps allow analysts to quickly identify which terms often co-occur, and related combinations of terms are clustered in one cell of the SOM.

This paper is structured as follows: In Section 2 we introduce the feature-based sentiment analysis techniques with new pixel cell-based sentiment calendars and geo-temporal map visualizations. Section 3 extends the feature-based sentiment analysis techniques to term associations with self-organizing term association maps. An evaluation of association measures and n-ary associations and their strengths and weaknesses are also given in Section 3. Section 4 presents two use cases with real-world movie Twitter data and web survey data (from July, 2007 to June, 2011) to validate the effectiveness of our techniques, and Section 5 provides our conclusions.

2 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 like text streams, it cannot be accurately predicted which features will show up and therefore it is undesirable to use a predefined list of features. The analysis must 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 *feature-based sentiment* algorithms [1] to measure the sentiment value, as shown in Figure 2. In addition, we save all other content-bearing words, such as verbs and adjectives, for further processing steps.

Sometimes though, a certain feature may have different sentiments in different contexts. This includes both the semantic context and the geo-spatial context of a feature within a review. Within different topics the sentiment that people have on the same feature may vary. In addition, people in different world regions may also 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.

2.1 Geo Sentiment Map

Sentiment analysis of customer feedback is often a process that excludes geo-spatial information. The sentiment analysis process mainly focuses on how the customers liked or disliked an object and what attributes of the object the customers assessed. Only a few analyses focus on the spatial distribution of opinions and show the influence of the geographic locations towards the sentiment. But it is crucial to take the geographic locations into account as they may influence the sentiment analysis of customer feedback. In the marketing process, for example, it may be important to analyze why the people of a region did not like a certain attribute of 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 different variants of a product that are adapted to local demands.

But adding the geographic information of opinions to the analytical process makes things more complicated. As soon as we deal with the locations of human generated data, we encounter different data densities resulting from varying population distributions. The unequal distribution of data points is a challenge for displaying the 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 process. 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 data points seen while dealing with the problem of the varying densities that cause overlapping. To overcome the overlap problem, we apply a pixel placement algorithm, as shown in Figure 3, to the data set to avoid overlapping points. Our pixel placement algorithm replaces the overlapping points with a circle of points entered on the nearest free position. Figure 3 describes 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 twitter feeds for the Kung-Fu Panda movie. The color of each pixel (review) represents the sentiment value (red: negative; gray: neutral; green: red). A high density areas, such as in Los Angeles and Philippines, is identified by a circle with non-overlapping

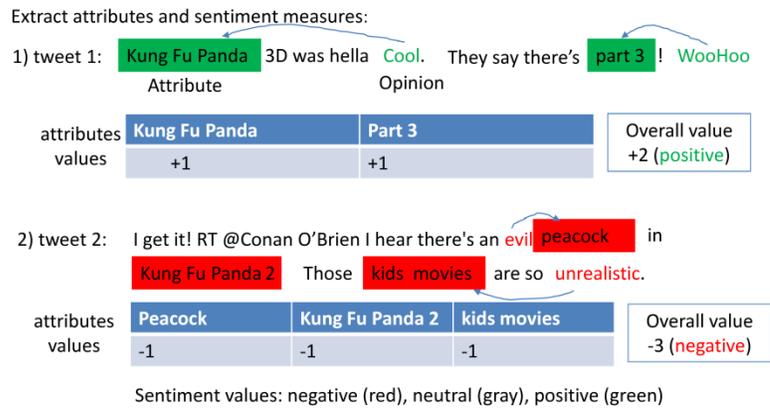


Figure 2: Methods to extract attributes and to measure sentiment values

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 {evil, peacock} appeared in a negative feedback from Los Angeles.

<pre>doPixelPlacement(OrderedList DataObjects) int[][] overlapCount := new int[width][height]; for each o of DataObjects do Point p := o.getPixelPos(); if (overlapCount[p.x][p.y] < maxOverlap) o.setPaintPos(p); overlapCount[p.x][p.y]++; else rearrangeDataObject(o, p, overlapCount); end for;</pre> <p>(a) Main method to replace all overlapping points</p>	<pre>rearrangeDataObject(o, p, overlapCount) int squareSize := getLastUsedSquareSize(p); Point[] squarePoints := calcSquarePoints(p, squareSize); while new place not found do if any squarePoints left Point p := next squarePoint; if (overlapCount[p.x][p.y] < maxOverlap) o.setPaintPos(p); overlapCount[p.x][p.y]++; else squareSize ++; squarePoints := calcSquarePoints(p, squareSize); end while; updateLastUsedSquareSize(p, squareSize);</pre> <p>(b) Helper method to reposition an overlapping point to the next free position</p>
--	--

Figure 3: 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 pixels location 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 3. 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 4 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.

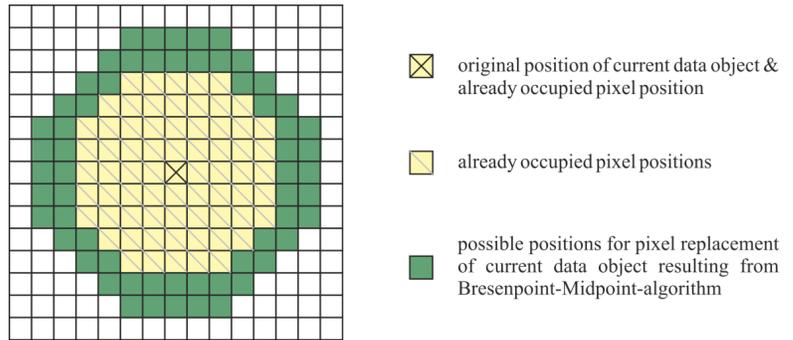


Figure 4: Schematic explanation of pixel placement process

2.2 Pixel Cell-Based Sentiment Calendar

Figure 1 (b) shows a pixel cell-based sentiment calendar arranged in a row and column format. Each row contains two categories: date and attribute (e.g., 5/29, panda or peacock). The columns are used as time intervals, such as hours, days, and month. Each attribute in a review (e.g., panda) is represented by a pixel cell. The cell color is the sentiment value (green: positive, gray: neutral, red: negative). Cells are arranged from bottom to top and left to right according to the customer review arrival sequence.

Using the above mechanism, we construct all columns (0-23 hours) in the calendar view as shown in Figure 1(b). Figure 5 shows two rows of three blocks for 5/29 at Hour 12, 13, and 14. Each block contains all the reviews for one attribute (e.g., panda) at that hour. The width of the block is computed from the window size and divided by the number of columns (e.g., 24). The height of each block varies according to the maximum number of reviews in the same row. The block at Hour 12 contains more reviews than those at Hour 13 and 14.

Customer feedback may occur at arbitrary points in time. In order to align all the feedback across different attributes in a calendar view, we introduce blank cells (gaps) to line up their arrival times. The reviews in “panda” are aligned with the reviews in “peacock” at Hours 12, 13 and 14 by using many blank cells to fill up the time distance. Therefore, all the blocks for the same day have the same width and height as the block that contains the largest number of reviews (Hour 12). This alignment method enables analysts to quickly identify patterns and uncover correlations from customer feedback streams in a same time interval.

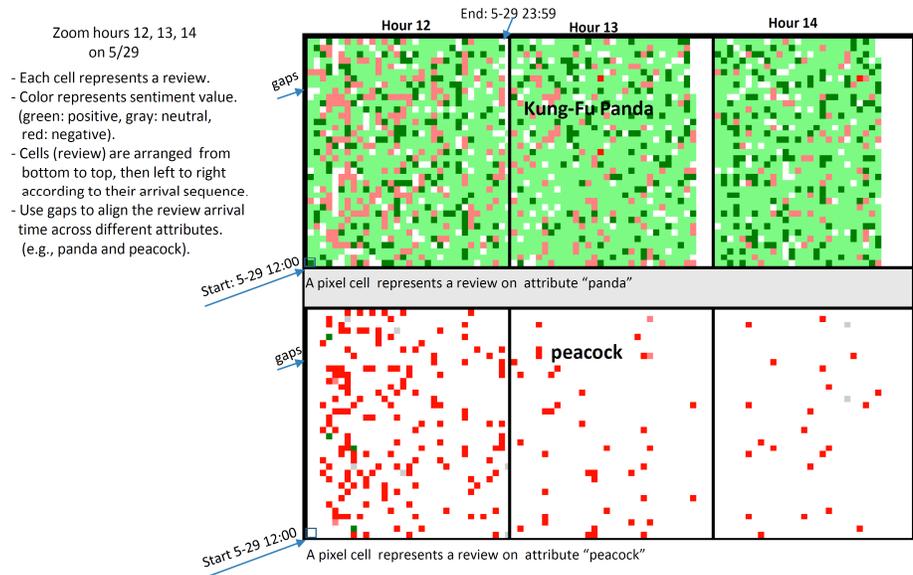


Figure 5: pixel cell (review) hourly placement in a calendar view, using gaps to align review's arrival time

3. Extend Sentiment Analysis to Term Associations

3.1 Term Associations

For a broader view the important terms and features, which are interesting to analysts, have to be brought into context. To this end, information about which terms are associated has to be automatically extracted from the text resources and visually conveyed to the analysts to enable them to gain a 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 has been seen, showing evidence that supports the rejection of the null hypothesis, the higher the level of confidence.

To apply such methods for term associations we first have to define the probabilities 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 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 before, the goal of extracting associations is to present them to a 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 for the web surveys contained {website, easy}, {website, to navigate}, and {easy, to navigate}. Evidently, these associations belong to the same statement 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 is 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 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 each, 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*. Sometimes, however, 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 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 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.

3.2 Self-Organizing Term Association Map (SOM)

Section 3.1 describes how individual words are grouped into associations based on their sentence-wise co-occurrence. One association represents one problem; for example, 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(c), 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 value. 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 also want to reflect how the clusters relate. Therefore, we generate a self-organizing term association map (SOM) with a simple square topology. In order not to overwhelm the analyst, the map is limited to 16 clusters, but it can easily be extended to cover more clusters or show a different topology. More details about the visual representation are given in Section 4.

3.3 Term Association Evaluation

Evaluation of Association Measures

It was not quite 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 (Φ). In order to get meaningful results we tested the methods on real data from web surveys. The data set consists of 47,966 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 6 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.

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.

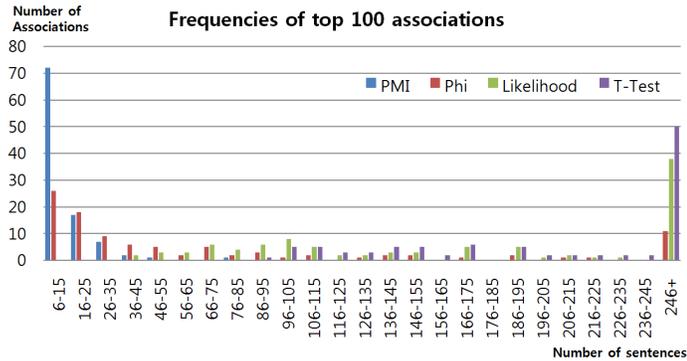


Figure 6: The frequency distribution of the top 100 associations extracted with each measure for the web surveys. The y-axis shows the frequency of associations in the corpus, i.e., in how many sentences associations occur. The x-axis reveals how many of the top 100 associations had a certain frequency.

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 the 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 just were 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.

4. Use Cases and Evaluations

Integration of sentiment analysis and term associations with the above visual analytics has a large number of applications, including hotel reservations, product surveys, IT services, Disney abstractions, movies, etc. To validate our approach, we have used two data streams (a movie and a web survey) to demonstrate the effectiveness of this integration.

4.1 Kung-Fu Panda Twitter Stream

Gain better spatial insights from geo sentiment map

We applied our geo-sentiment map, introduced in Figure 1(a), 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. Figure 7 illustrates the differences in both the volumes and sentiment values during the preview and the opening week. The opening week had 60 times more reviews than the preview. Analysts can quickly identify the differences from the geo-location patterns.

To evaluate the effectiveness of this geo sentiment map, we compare it (Figure 8(b)) with the ordinary map shown in Figure 8(a). In Figure 8(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 8(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 regions or countries.

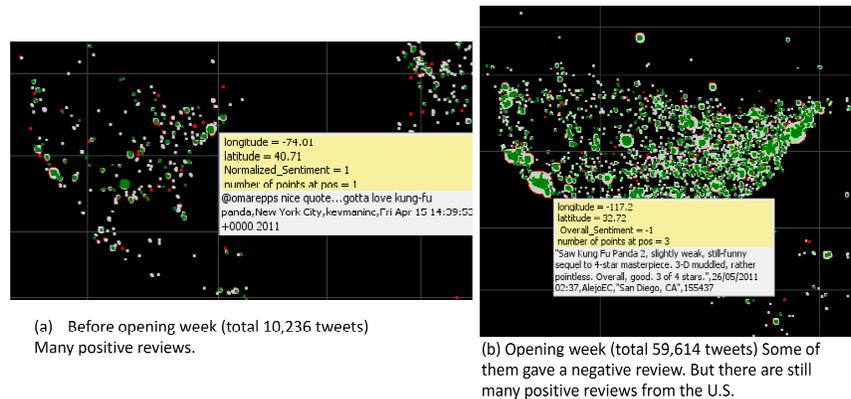


Figure 7: Visual comparison between Kung-Fu Panda previews and opening week using geo-location patterns in the US map. There are more high density locations during the opening week than during the previews. Many people gave positive comments.

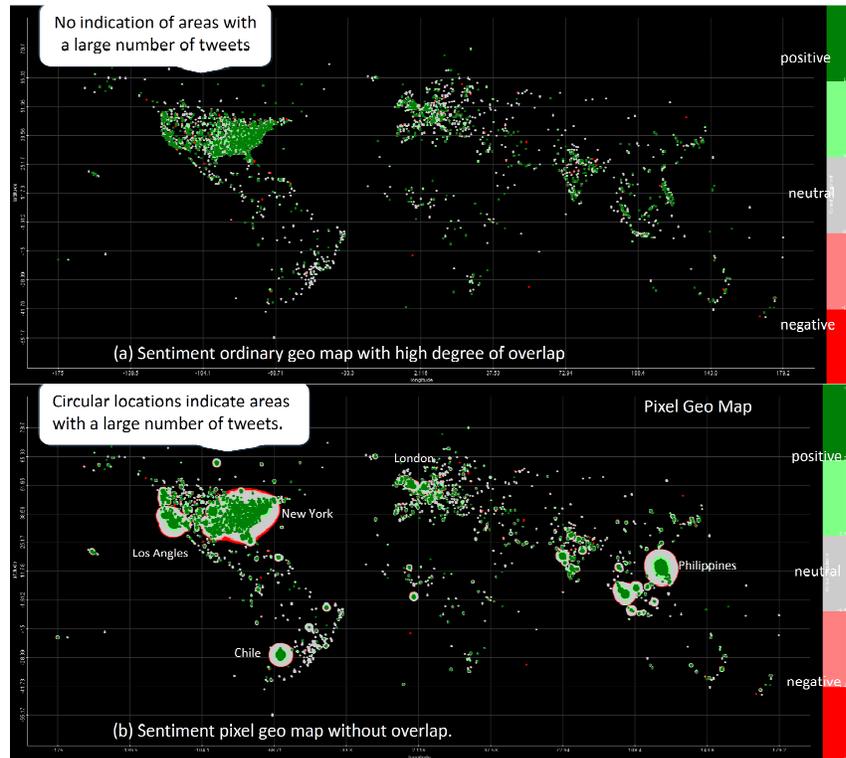
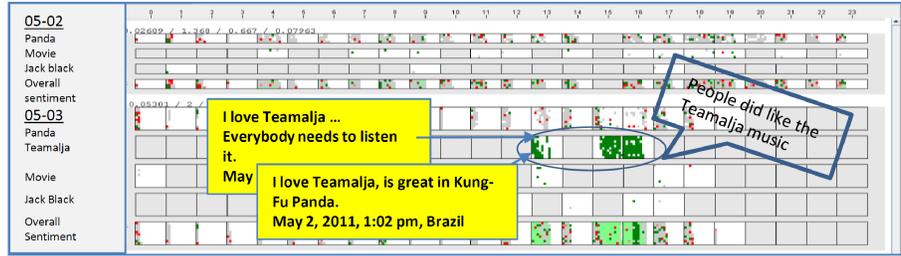


Figure 8: Geo map high density area evaluation (i.e., New York and Los Angeles)

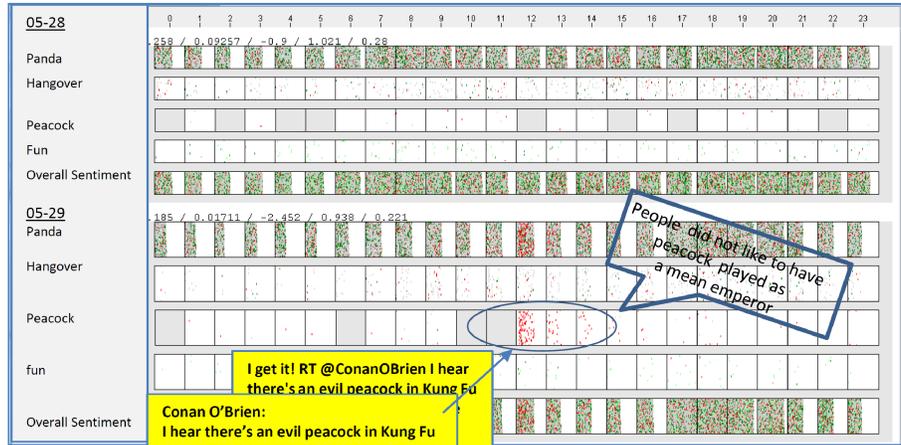
Gain better temporal insights from pixel cell-based sentiment calendar

Figure 9 shows two different calendar views. The top calendar is generated from the tweets during preview time and the bottom calendar is generated during the opening week. Each review is shown as a pixel (cell). The color is the sentiment value. Each calendar has some interesting rows corresponding to term occurrences such as Panda, Teamalja, Jack Black in the preview, and Panda, Peacock, fun in the opening week. From both calendars, analysts can quickly identify the temporal patterns from both the preview and the opening week by the following facts:

- There are very few reviews for the preview. But each day the number of reviews grew (more pixel cells). Most of the audience gave their reviews in the afternoon and early evening (hours 13 to 19).
- For the opening week, comments on Kung-Fu Panda increased from 10,236 reviews to 59,614 reviews from all over the world. The increase in the number of reviews did not impact the sentiment calendar view. Analysts can easily analyze the opening week sentiment results without clutter in the display.
- The most popular attributes commented on are Panda, Hangover, Peacock, fun etc. Most of the reviews are favorable to Panda over Hangover (more green reviews).



Kung-Fu Panda Preview



Kung-Fu Panda Opening Week

Figure 9: Customer reviews for the preview and the opening week. (row: hours, column: date and attribute list, color: the sentiment value. Each dot is a review. Dots are ordered from bottom to top, and then left to right

The increase in the number of reviews did not impact the sentiment calendar view. Analysts can easily analyze the opening week sentiment results without clutter in the display.

There are two interesting observations on 05-03, positive reviews increased suddenly for the Ku-Fung Panda music, Teamalja, due to some influential reviews that had been previously posted. Then on 05/29, a large number of negative reviews on peacock were sent seconds after one specific negative review in hour 12. After drilling down on the first negative review, the analysts discovered that the other negative reviews were influenced by the negative comment of a famous person, Conan O'Brien, thus validating our approach.

Gain better insights from term association using sentiment self-organizing term association map

As illustrated in Figure 1(c), users can quickly identify attributes, verbs, and adjectives that frequently occur together. For example, "Panda" frequently associates with "awesome". Therefore, the sentiment geo map shows that the majority data points (reviews) are positive. The surrounding areas for Los Angeles show some red due to the influence of the negative review.

4.2 Web Survey Data Streams

Geo Sentiment Map

As illustrated in Figure 10 for web surveys the population density overshadows other aspects. For example, New York City and Los Angeles have dense populations, and hence are likely to produce many comments. However, it may be that the experience in a less densely populated state in Alaska is significantly different than that in Los Angeles. For example, the delivery experience in Alaska is

different, because it is so lightly populated. It is more important for them not to have to go to the nearest store to buy ink, since the nearest store may be 4 hours away. Hence, extracting associations for particular regions may provide some insight into regional differences. The future challenge will be to spot patterns beyond population density.

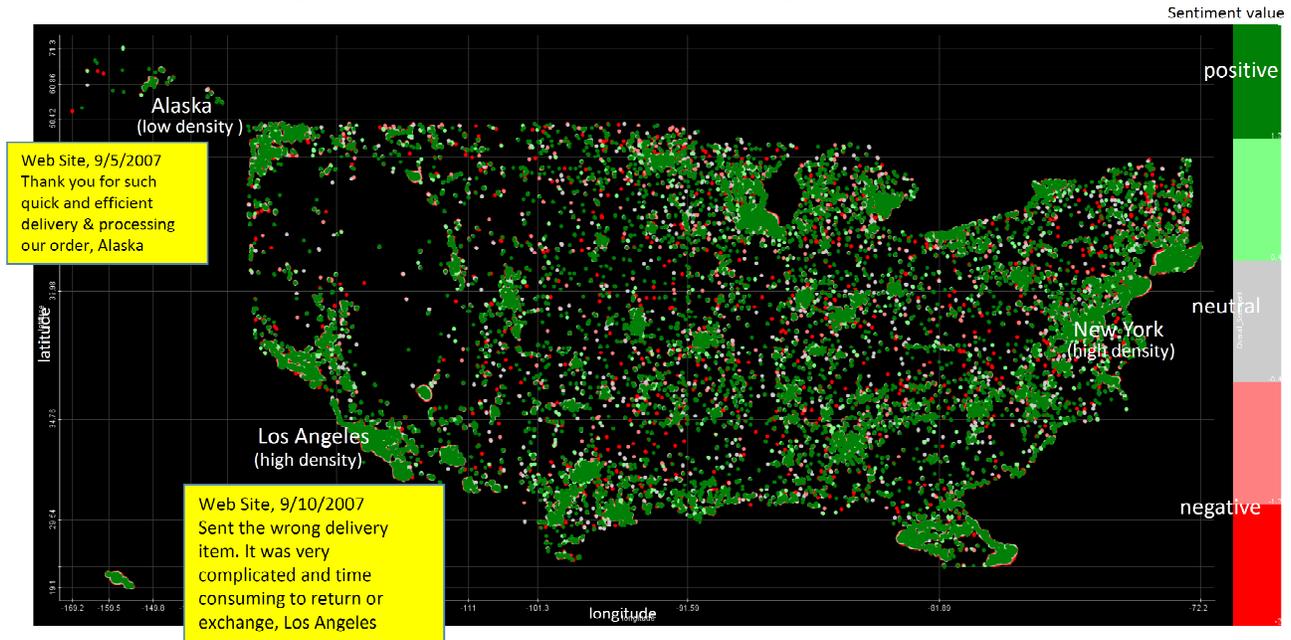


Figure 10: A buyer’s web survey pixel geo map (2007-2010, 52,189 customer feedback comments)
 The color map indicates the customer feedback on a buyer’s web survey (red: negative; gray: neutral; green: positive) and shows that the population density overshadows other aspects.

Cell-Based Sentiment Calendar

Figure 11 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. Service managers can quickly observe the variances e. g., printer and website have more red than delivery and shipping. Service 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 11(b)) and other attributes which have a high correlation with “printer” (Figure 11(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 & is noisy”. This observation validates that the terms {printer, ink} often occurred together.

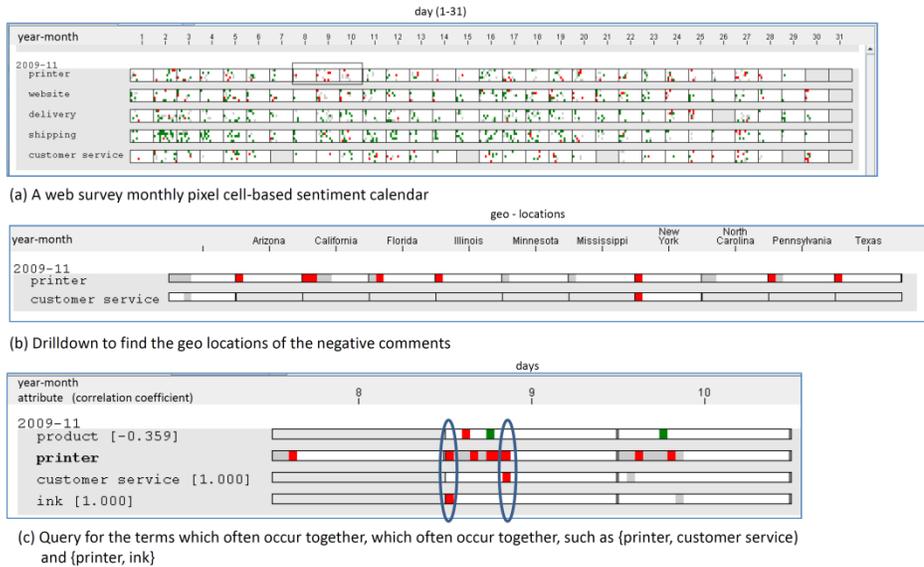


Figure 11: (a) A web survey (buyers) monthly pixel cell-based sentiment calendar
 (b) Users are able to rubber-band around 9/8/2009 and drilldown to find the geo locations of the negative “printer” feedback.
 (c) Users can issue a query to locate the terms associated with attribute “printer”.

“Printer support is great but this printer gobbles ink & is noisy”. This observation validates that the terms {printer, ink} often occurred together.

Self-Organizing Term Association Map

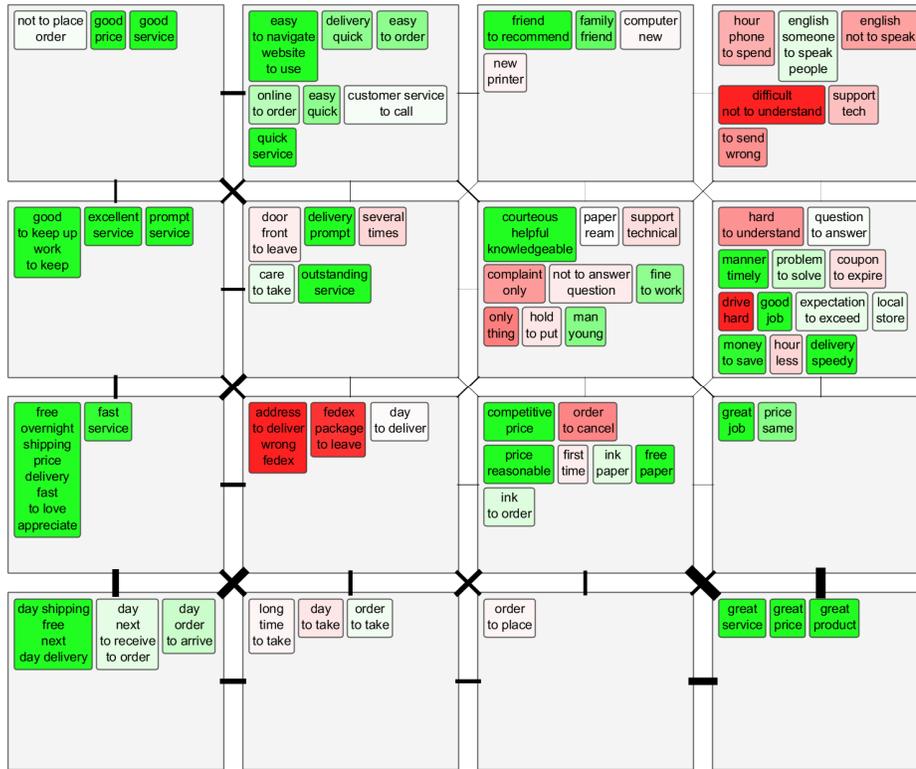


Figure 12: A clustering of all associations for the buyer's web surveys into a self-organizing map with 16 nodes

In order to explore whether our visual analysis interface reveals valuable additional information to analysts, we generated the corresponding self-organized map (SOM) in Figure 12. The SOM-nodes are displayed as rectangles and connected with lines. The thicker the line between two SOM-nodes, the closer the respective cluster centroids are in the underlying high-dimensional vector space. Based on this visualization we performed a case analysis with real data from the buyer web survey described above. At first sight it becomes evident that there are many more positive comments than negative ones.

With respect to negative associations two clusters are dominant. The cluster on the top right deals with problems regarding the language skills of the customer support teams, who some customers find difficult to understand. We also see that some customers complain that they have to spend hours on the phone solving one problem with customer service. The other cluster (further down to the left) deals with the order delivery experience. Some people complain about FedEx delivering their order to the wrong address and some complain about FedEx leaving the package in front of the door. When the user moves the mouse over an association, the respective sentences from the underlying text data are displayed in a tooltip. This interaction reveals the problems that people have with FedEx leaving a package. Some are afraid that it could have been stolen and others complain about rain damage.

The dominant positive feedback is easily analyzed. First of all, people love the good service and especially the free shipping and one-day delivery. Also, the competitive prices and free paper are appreciated, as well as the fact that the website is quick and easy to use for ordering online. Finally, people state that they would recommend the company to their family and friends.

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". A further interesting peculiarity is that the cluster on the upper left contains the associations {good, price} and {good, service}, and the cluster on the lower right contains {great, price} and {great, service}. These statements appear to be very similar, but are not clustered together. This is due to the fact that the persons who wrote "good price" never wrote "great price" in the same document and vice versa. Right now, two associations are clustered together if they appear

together. To collapse the distinction between “good price” and “great price”, we have to cluster associations that appear in similar contexts but not necessarily within the same documents. In conclusion, this first visualization allows us not only a surprisingly detailed insight into the document collection, but also points to the improvement potential of the underlying algorithmic processing, thus serving an important purpose of visual analytics.

5. Conclusion

With the current high speed, high volume of customer feedback streams, new sentiment techniques are required for helping companies to know what their customers like or dislike about their products and services in real-time. In this paper, we presented a novel integration method that comprises 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., attributes, verbs, and adjectives) which occurred frequently together. Our integrated analysis system extends the scope of the sentiment information which users need to know about their customers. In visualizing such a large volume of feedback, there were two main issues: scalability and density. To solve both problems, we introduced pixel sentiment calendar and pixel geo-map. Using a pixel sentiment calendar, analysts can gain better insights into temporal patterns over a large customer feedback stream. With a pixel geo-map, analysts can gain better insights from sentiment geographical distributions and are able to quickly identify the high density areas. From our experiments, population density overshadows sentiment aspects. To visualize hundreds of terms in a single view, we introduce a variant of a self-organizing map. We cluster related terms into different nodes. The color of a term represents the sentiment value in a sentence, which is the result of summing up all the sentiment words contained in the sentence. From the sentiment value, analysts can quickly identify the important terms to perform proper actions.

The above combined techniques 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 provided corresponding improvements and suggestions. Our future work will proceed to detect time-related sentiment patterns, trends, and influences in the customer feedback streams for live alerts.

ACKNOWLEDGMENTS

The authors wish to thank Laura Hill for her suggestions and encouragement, and Malu Castellanos and Riddhiman Ghosh for providing Kung-Fu Panda customer feedback stream, 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.