A Survey on Sentiment Analysis and Visualization

Ameni Boumaiza
College of Engineering, University of QATAR
Email: boumaiza.amani@gmail.com

***Abstract*—The growing popularity of sentiment visualization systems in recent years is an active research area. In this paper, we will present a survey that reviews and assesses the recent visualization techniques and systems in the field. This report classifies and reviews the recent approaches in visual analysis. The motivations in conducting this survey are twofold. First, we aim to review the most recent research trends and developments of sentiment visualization techniques and systems and provide a precise review of the field. Second, this survey aims to provide a critical assessment of the research which can help enhance the understanding of the field.**

***Index Terms*—sentiment analysis, visualization, third term, fourth term, fifth term, sixth term**

I. Introduction

Online Social Networks become the medium for a plethora of applications such as targeted advertising and recommendation services, collaborative filtering, behavior modeling and prediction, analysis and identification of aggressive behavior, bullying and stalking, cultural trend monitoring, epidemic studies, crowd mood reading and tracking, revelation of terrorist networks, even political deliberation. They mainly aim to promote human interaction on the Web, assist community creation, and facilitate the sharing of ideas, opinions and content. Social Networking Analysis Research has lately focused on major On-line Social Networks like Facebook, Twitter and Digg [1]. However, research in Social Networks [2] has extracted underlying and often hidden social structures [3] from email communications [4], structural link analysis of web blogs and personal home pages [5] or recently explicit FOAF networks [6], structural link analysis of bookmarks, tags or resources in general [7], co-occurrence of names [8] [7], and co-authorship in scientific publications references [9], and co-appearance in movies or music productions [10]. Interactive visualizations is employed by visual analytics in order to integrate users' knowledge and inference capability into numerical/algorithmic data analysis processes. It is an active research field that has applications in many sectors, such as security, finance, and business. The growing popularity of visual analytics in recent years creates the need for a broad survey that reviews and assesses the recent developments in the field. This paper reviews the state of the art of sentiment visualization field. Opinion mining or sentiment analysis refers to the application of natural language processing, text analytics and computational linguistics to identify and extract subjective information in source materials (Source: Wikipedia). Sentiment analysis is a multi-disciplinary artificial intelligence problem and aims to classify user sentiments, positive, negative and neutral. Opinion mining generates a list of features and aggregates opinions about each feature. It aims to summarize and visualize people's opinion about products, topics, and services from online reviews. In literature the terms opinion and sentiment are of-ten used interchangeably. In our paper we will use these terms also interchangeably. Sentiment analysis systems are being applied in almost every business and social domain because opinions are central to almost all human activities and are key influencers of our behaviors. There has been a massive increase on the use of web technologies which have caused an increasing number of social networking platforms used such as forums, blogs and micro blogs. The information in such blogs are multidimensional, time varying and mutable that's why it is difficult to analyze and visualize the features of popular topics. To address these challenges, sentiment visualization techniques and systems has been developed in recent years. A combination between opinion analysis and interactive visualization was proposed in order to involve user into the data analysis processes evaluation. Many platforms were developed for users to share the opinions and views about the products, topics, events or services.

This paper describes the state of the art visualizations methods that are tightly coupled with the goal to enable users to detect interesting opinions from social media. The interestingness is derived from the sentiment, temporal density and context coherence that comments about features for different targets (e.g. users, companies, product characteristics, topics, etc.) have. Visual analytics contributions are made at different stage including novel ways to visualize sentiments, opinions for further exploration. Sentiment visualization task aims to detect and pre-select interesting time interval patterns for different features in order to guide analysts. Business analysts are the main target group who want to explore time-stamped customer feedback to detect critical issues.

Manuscript received October 14, 2015; revised December 14, 2015.

This state-of-the-art report reviewed recent re-search in the field of visual analytics. It represents a comprehensive overview of many advances in visual analytics techniques and applications to gain a better understanding of the cutting-edge research in the field. Evaluating visualization tools and techniques confronts researchers with many questions. The next section will describe a survey which is focused on some state-of-the-art visual analysis trends and visualization techniques. The objective of this study is to highlight existing works in this field, to present the characteristics of each techniques, its level of abstraction, advantages and disadvantages.

II. Opinion and Sentiment Analysis

Social network revolution plays a crucial role in gathering information containing public opinion. To obtain subjective and factual information from this in-formation, public opinions are extracted [11]. Thus, it is the process to predict hidden information about user's positive, negative or neutral sentiment about a topic. These social networking sites generate enormous data up to tera bytes per week. Many techniques have been proposed to characterize the extracted information from social media. These approaches have proven their usefulness in many applications. Information visualization generally aims at providing new insights and helps understand some structure of the data. There are many ways to analyze and visualize information. Much of the work in this area has focused on topic-based methods and feature-based analysis.

A. *Topic-based Methods*

Topic-based methods extract topics or events from text corpora and visually explore the extracted information using different techniques. These methods involve the use of the tree map and visual summary report. In this representation, two dimensions of information (topic and sentiment) need to be dis-played simultaneously. The user can easily access the specifics of a given topic. Recent researches, such as EventRiver [12], Visual Backchannel [13], and TextFlow [14], mostly analyze and track on the temporal evolution and diffusion of events, topics, or activities.

Pulse [15] is a system for mining customers' reviews on cars. It is a Tree Map-style user interface which was designed to visualize the topics and their sentiment values. It utilizes a Tree Map visualization [16] to display clusters and their associated sentiment. Each cluster is rendered as one box in the Tree Map. The taxonomy of makes and models (i.e. major and minor category) is displayed in the left pane, the Tree Map to the right of it, and the sentences in the tabbed display at the bottom [15]. The size of the box indicates the number of sentences in the cluster, and the color indicates the average sentiment of the sentences in the box. The color ranges from red to green, with red indicating negative clusters and green indicating positive ones. Clusters containing an equal mixture of positive and negative sentiment or containing mostly sentences classified as belonging to the "other" category are colored white. Each box is also labeled with the key word for that particular cluster.

In Opinion wheel [17] and rose plot [18] systems, data are arranged in an elliptical or circular way. The opinion wheel used to present the hotel customers' feedback data, utilizes a combination be-tween an opinion triangle, which consists of scatter plot, and an opinion ring. The vertices of opinion triangle represent positive (P), negative (N) and un-certain (U) opinions. Each Point inside the triangle represents opinions and their positions encode se-mantic orientation. Categories of different data dimensions are represented by the opinion rings, with colored histograms.

In [19], authors used a bar chart as a visualization technique. Each bar above the X-axis gives the number of positive opinions on a feature, and the bar below the X-axis gives the number of negative opinions on the same feature.

A visual comparison of consumer opinions on two competing digital cameras. Consumers reviews about the features of each camera can be clearly seen. [20] created a bar chart-based trend visualization to analyze the RSS data of US presidential election in 2008. Visual analysis of graphs is an important application of visual analytics. Interested readers can refer to a complete survey [21] for more details about the past research. The graphs adapted for visualizations in opinion mining include coordinated graph [22], comparative relation map [23], positioning map [24], line graph and pie chart [25]. [22] introduced a novel visualization approach for visual analysis of how positive and negative reviews of the controversial bestseller The Da Vinci Code differ. A graph composed by the selected terms as nodes (blue for positive, red for negative, magenta for mixed) is used in this approach. Node size encodes total appearances of each term. Edges connect terms that appear the same month, with thickness representing the number of months in common. Selected terms are highlighted in red in the arc diagram. If the time filter checkbox is selected, the tables and graph filter out terms for months outside the time range visible in the arc diagram. Analysts can brush interesting terms in any of the views, explore patterns of term usage by panning and zooming over time, then drill down to compare temporal patterns for particular terms. The top and the bottom half of the graph present terms used in positive and negative reviews, respectively. The vertical bar thickness shows the number of terms for each month. The arcs connect months in which common terms appear and their thicknesses represent the number of common terms. The authors also used a spectrum graph to display the distribution of positive and negative reviews over time. The coordinated graph provides an effective way for the comparison of terms with respect to time. Similarly, Xu *et al*. [23] used graph to compare NokiaE71 with its competitors, that is, Nokia E61, Blackberry Curve, iPhone, and Blackberry Bold 9000, on different features, such as size, function, and looks. The red box shows the number of favorable statements for Nokia E71, and the blue box shows the number of favorable statements for competitive products. The comparative relation map was very helpful (1) to highlight the relative strengths and weakness of products, (2) to analyze threats from competitors and enterprise risks, (3) to support decision making and risk management, and (4) to design new products and marketing strategies. Many others examples of the use of graphs for opinion mining visualizations that utilize positioning map, line graph, and pie charts. [24] used two-dimensional positioning maps for the comparison of competing cellular phones. Five cellular phones with their corresponding four characteristics (keywords) were compared and plotted on a map. The top four characteristic words for each phone are plotted around each cellular phone. We remark that cellular phone A is well reputed for its basic performance ("fast", "no Problem") instead of the cellular phone B which shows a bad reputation ("doesn’t work well", "slow "). Finally, [25] developed and presented a visualization system Amazing for mining useful knowledge from consumer product reviews. They introduced a line graph to present the number of positive and negative comments about a product over time, and a pie chart to show the proportion of positive and negative reviews. This section reviewed recent research of topic-based methods. Although we have witnessed rapid developments in the research area, it is still very difficult to visually analyze and explore more works.

III. Feature-based Sentiment Analysis

Most approaches for feature-based sentiment analysis involve three or four consecutive steps [26]:

- Features for different targets (e.g. persons, companies, products, services or topics) are detected either directly from the corpus or based on predefined word lists.
- Sentiment words that describe the extracted features are searched in the documents. Sentiment words are words that evoke positive or negative associations.
- A mapping strategy aims to detect which sentiment words refer to which feature, so that a sentiment score can be determined for each feature.
- Some approaches visualize the results of the feature-based sentiment analysis and enable the user to interactively explore the results in detail.

For the first two steps abundant research has been published in the last years. For the sake of brevity we refer to comprehensive summaries given in [27] and [28] for details. Both features and sentiment words can be either learned from the processed text documents itself, from external resources (like e.g. WordNet1) or they can be gathered from predefined lists. One special challenge is to identify sentiment words that have no general validity, but depend on the domain or even on the feature. Feature-based sentiment analysis is a subtask of opinion and sentiment analysis. Feature-based methods use various features such as word-level features [29] and document-level features [30] to visualize text. Word clouds have received a great deal of attention and are commonly used in the last few years. This method provides an intuitive visual summary of document collections by displaying the keywords in a compact layout. Key-words that appear more frequent in the source text are drawn larger. A variety of algorithms such as Wordle [29] and ManiWordle [31] have been proposed to create good word cloud layouts. However, the semantic relationships between the keywords in the original text are lost in the layouts. To handle [32], [33] used various visualization techniques such as: visual summary reports, cluster analysis, and circular correlation map to facilitate visual analysis of customer feedback data at the feature level. In [32], authors described an interesting visual analysis application for answering "How to make your writings easier to read". Their work uses a semi-automatic method to choose proper features from 141 candidate readability features. They developed a visual analysis system called VisRA with three views, including the corpus view, the block view, and the detail view, to explore the feature values of text corpora at different levels of detail. Researchers introduced methods such as the force-based algorithm [34] and the seam-carving algorithm [35] to produce semantic-preserving word clouds. This can ensure that the keywords that co-occur frequently in the source text are placed close to one another in the word clouds. Sentiment-to-Feature Mapping In order to deter-mine which sentiment words refer to which feature, different approaches have been suggested in the past. Distance-based heuristics is used by some approaches, i.e. the closer a sentiment word is to a feature word, the higher is its sentiment influence on the feature. [36] operates on whole sentences, [37], [38] operate on sentence segments, or on pre-defined word windows in [39]. Advanced natural language processing methods are exploited such us typed dependency parsers, to resolve linguistic references from sentiment words to features. [40] use dependency relations to extract both features (product attributes) and sentiment adjectives from reviews by a double propagation method. [41] use subject-verb, verb-object and adjective-noun relations for polarity classification. Pairs (sentiment word, feature) are extracted and based on 10 extraction rules that work on dependency relations. [42] use lexico-syntactic patterns in a bootstrapping approach for subjectivity classification resolving relations be-tween opinion holders and verbs. Another approach [17] was recently published that takes uncertainty into account. The authors consider if customers do not express clear opinions, i.e. "customers" conflict and uncertainty about their opinions as well as the uncertainty involved in the automatic opinion analysis processing. Thus a feature mention can be both negative and positive at the same time. Their uncertainty score is based on two parameters: the smaller the difference between the negative and positive sentiment on a feature within a sentence and the longer the sentence, the higher the uncertainty [26].

This section mainly introduced recent topic-based and feature-based methods for visual analysis of text data. Both kinds of methods are commonly used to solve practical problems. Visual analytics of text data still faces a few challenges although some successes that have been achieved. It is still difficult for current methods to handle large amounts of text data. To overcome this limitation, more efficient text mining and NLP algorithms as well as scalable interactive visualizations, are needed. Another challenge is to handle the natural language ambiguity and the un-certainty that arises from the text mining algorithms. Finally, text data is often accompanied by multimedia data such as images and videos, which are even more challenging for analysis. Heterogeneous text data with images and videos can be complementary. It may allow users to explore the data from different perspectives. Thus, effective analysis of the heterogeneous text data is worth further study [30]. In order to allow users to visualize the outcome of automatic feature-based sentiment analyses and enable further user explorations, several different approaches have been suggested. The next section describes relevant related work on visual topic-based text visualization and topic-based social network visualization.

IV. Visualization Systems

Information visualization generally aims at providing new insights and helps understand some structure of the data. The objective of opinion visualization systems is to offer a simple way to end users to reconstruct meaningful summaries of reviews or events.

Sentiment visualization is a subfield of it that deals with the presentation and visual analysis of opinions extracted from social media. There are many ways to analyze and visualize sentiment. Much of the work in this area has focused on analyzing users's sentiment. Different opinion visualization systems have been proposed in the literature. However, few studies have been done on finding effective means to present users opinions with sufficient visual ways and different levels of summarization. We perform this literature review in order to investigate features and methods, analyze state of the art visualization systems, provide comparison between these systems and provide useful conclusions.

A. Topic-based Text Visualization

Topic-based text visualization has become a widely researched topic over the past few years [43, 14, 44, 45, 46, 12]. It has received considerable attention of researchers. Numerous methods [44, 47, 48] leverage a river metaphor to convey evolving topics over time. For example, an information visualization system called Theme River [44] visually depicts the changes in keyword strength over time using a river metaphor within a large collection of documents. A new visualization method called TextFlow [14], enables analysts to visually analyze topic merging and splitting relationships over time using Sankey diagrams and graphs. It models the relationships between evolving topics, assists analysts in the visual analysis of topic merging/splitting relationships and tracks their evolution over time. [49] explored temporal changes of topics and utilized the topic rose tree algorithm to construct a topic hierarchy in large corpora and the hierarchical Theme river view. [47] and [48] combined the river metaphor with other visualizations to study the spreading patterns of topic and sentiment in social media. Interactive visualization with topic analysis techniques are integrated by TIARA [46], [50] to assist users in understanding a document collection. It employs the LDA model [5] to analyze a large text corpus and visually illustrates the evolution of topics using an enhanced stacked graph. [51] developed Meme Tracker to effectively identify a huge number of topics from millions of news articles. A coherent representation of the news cycle in thus provided, allowing users to track the temporal behaviors of memes represented by short phrases in blogs and news.

Recently, several visualization techniques were proposed to help users analyze temporal events and their evolving patterns. Event River [52] assumes that clusters of news articles with similar content are adjacent in time and can be mapped to events. Based on this assumption, this method automatically detects and presents interesting events to reveal their impact over time. Life Flow [53] and Outflow [54] facilitate the exploration of temporal event sequences by aggregating multiple event sequences into a tree and a directed acyclic graph, respectively. Afterwards, a timeline visualization is used to display the aggregated event sequences from multiple aspects. A set of user-driven data implication techniques was developed by [55] in order to help users understand large temporal event records. The above approaches focus on the visual exploration of evolving topics/events with flat structures. A set of heuristic rules was developed by [56] to quickly generate a storyline layout. Although this method can generate a storyline layout in real-time, it may result in a layout with many line crossings and wiggles. To overcome this limitation, [57] formulate the storyline layout as an optimization problem. They choose to leverage spatial information in order to convey hierarchical relationships among characters. However, hierarchical information is not considered in the layout algorithm and is added in the post-processing step instead. As a result, it cannot leverage the location hierarchy to handle a large number of character lines. An efficient optimization approach was proposed by [46] to support real-time interaction. They formulate the storyline layout as a novel hybrid optimization approach that combined discrete and continuous optimization. It utilizes a predefined location hierarchy to handle thousands of entities in the storyline layout. Hierarcical Topics [49] and Topic Panorama [58] employ the BRT model [59] to construct a static topic hierarchy from a set of topics, which is used to present changes in the topic content in a hierarchal manner. Although the aforementioned methods can manage scalability by utilizing a static hierarchy, they may fail to concisely reflect the new topic structure as new documents appear [60]. In order to overcome this problem, [43] propose a method which differs from these approaches as it generates a set of evolving topic trees, which are used to smoothly organize a large number of topics over time. To better present evolving hierarchical topics on limited screen real estate, they propose an incremental evolving tree cut algorithm to extract an appropriate number of topics for each tree based on the user selected focus node(s).

B. Visual Opinion Analysis

In addition to topic-based text visualization applications, some works were focused on social networks visualization and visual opinion analysis that has been the subject of increasing attention for re-searchers. A system called Mood Views was presented by [61]. The system tracks the mood of blogs hosted by Live Journal. Conceptually similar to [62], the system continuously downloads updates from thousands of blogs. It tracks moods, predicts what they might become and analyses them in an attempt to understand why specific changes in mood occur. The mood tracker follows moods in close to real time and creates graphs based on these time series. Natural language processing and machine learning are combined by the mood predictor in order to estimate moods based purely on textual content. Mood data gathered from tagging and an accuracy statistic can be determined by comparing the estimated mood to the actual. It finally identifies terms which occur more often or less often than usual during certain peaks in mood using language statistics. In [19], a bar shows the polarity related with each product and each feature. The portions of the bar above and be-low a horizontal line represent the amount of positive and negative reviews. The UI (user interface) of [19] enables analysts and users to interact with the system. [17] proposed a system called Opinion Seer which has two major views: An opinion wheel (integrates a scatterplot with a radial visualization) and tag clouds. In addition to positive and negative values, the extracted opinions also explicitly contain the uncertainty values to indicate the amount of ambiguous information. It provides a set of rich user interactions. Aside from basic interactions such as pan and zoom, and some special user interactions for the system: brushing, moving projection center, area preserving mapping and opinion grouping. Although this system has proven its usefulness, it presents some gaps. Cultural background is not taken into account as a variable or uncertainty and authors considered the negative words to indicate the uncertainty of customer opinion (positive and negative words often indicate the uncertainty of customer opinion). They don’t include the spatial aspect in the visualization. [20] described a concise visual encoding scheme to represent at-tributes such as the emotional trend of each RSS news item. The novelty resides on the combination of a sentiment analysis method with a visualization technique revealing the emotional content of RSS news feeds over time. Several possibilities to interact with the tool: zooming, details on demand, similarity search and filtering. The system suffers from some limitations. We remark that only simple documents similarity measures are used and the presented analysis tool is not based on sophisticated sentiment analysis methods. [22] presented a visual analysis system using multiple coordinate views such as decision trees and terminology variation to help users to understand the dynamics of conflicting opinions. Both approaches [22, 20] for analyzing text contents are efficient by using word matching methods and they lack semantic analysis. [63] proposed an interactive visualization system called Senti View that aims to analyze public sentiments for popular topics on the Internet. The interactivity with users is based on using the notion of helix with attribute astrolabe and relationship map, it supports interactive visualizations of the time-varying sentiments of participants. Using this system, it is still very difficult to visually analyze and explore the factors influencing positive or negative opinions. [39] proposed an approach to automatically analyze large volumes of customer reviews. The system uses visual summary reports, cluster analysis, and circular correlation map to analyze customer feedback. [64] used time density plots and a pixel map calendar to allow feature-based opinion exploration of text document streams. Opinion Blocks [65] creates a feature-based visual summary of opinions from user reviews with a combination of advanced opinion mining techniques and crowd sourcing. Feature-based opinion visualization is particularly useful for exploring regular customer reviews, but it may not work for analyzing opinions from short microblog messages that usually do not have features. Scatterplots [66, 24], bar charts [67, 20], and rose plots [67] have also been used to visualize document-level opinions from review comments. The Pulse system [15] uses a tree map to display the topic clusters and their opinions. [62] proposed an interactive Web-based tree map, News map, to represent the relative number of articles per news item. [68] used line graphs indicating term trends in order to find the evolution of topical trends in social media. These works are focused on social networks, text analysis and knowledge representation of social networks to analyze microblog and forum content without sentiment analysis. [69] used to cluster keywords into themes and tracking their temporal evolution in order to repersent the change of stories. [70] developed an interactive visualization system to allow users to visually construct queries and view the results in real time. [67] proposed for sentiment mining and analysis, a user directed sentiment analysis method to visualize emotion, they only use statistical methods. Visual opinion analysis of microblog messages has attracted much attention from the field [46, 71] in recent years. TwitInfo [72] recall-normalizes aggregate opinion in-formation to produce trustworthy opinion overviews through pie charts. [73] present a visual analysis system with a pixel opinion geo map, key term geo map, and self-organizing term association map to facilitate exploration and analysis of customer feed-back streams. [74] proposed two different opinion dynamics models using the aforementioned theory [74]. [75] described how visualizations about the evolution of events on twitter are created by presenting several case studies in recent years. [13] provided an interactive multi-faceted visual overview of large-scale ongoing conversations on twitter, including a spiral to present participants and their activities and an image cloud to enclode the popularity of event photos by size. [76] presented SentiFul to automatically generate and score a new sentiment lexicon. [77] used the JST model that is based on LDA in order to detect sentiments and topics simultaneously from text. [78] analyzed the sentiment of restaurant reviews in Cantonese using classification. These approaches provide analysis and visualization on a single sentiment aspect. [79] proposed a method to show real-time information diffusion from Twitter using multiple viewing options. [80] developed a system called Mood Views, which is a collection of tools for analyzing, tracking and visualizing moods and mood changes in blogs posted by Live Journal users. Similarly to Theme River, [81] focus on the visualization of both topical and sentiment flow along within a timeline. The method accepts texts with time stamps such as Weblogs and news articles. Two kinds of graphs are produced:

Topic graph: shows temporal change of topics associated with a sentiment

Sentiment graph: shows temporal change of sentiments associated with a topic

This approach is the more common way to illustrate temporal opinion mining results; having a specific topic or event, and seeing how sentiment toward it changes and fluctuates as time goes by. Combining the analysis of these two types of graphs, the result is a solid visual representation of the association between sentiment and event. However, this approach requires both sentiment and event to be known prior to analysis. It does not attempt to dis-cover and identify previously unknown events which might have had an effect on sentiment. TwitInfo, a prototype system for monitoring events on Twitter, uses a timeline graph showing the major peaks of publication of tweets about a particular topic, the most relevant tweets, and the polarity of the opinions they express [72]. [82] developed three inter-active widgets that are arranged together to create coordinated multiple views: the temporal sentiment dynamics and sentiment comparisons among different categories/topics are visualized in the sentiment trend view, the associated structured facets are illustrated in the chart visualization view, and the de-tails of documents and context of sentiment are pro-vided in the snippet/document panel. The user is able to navigate the data set using optimal interactions due to the mash-up capabilities among the three views. Much related work exists on analyzing twitter feeds. Bifet and Frank [83] 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. [72] 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, [73] use feature-based sentiment analysis [33] with multi-resolution high density techniques to process large customer feedback stream in real-time. Authors then analyze each feature to see if it is mentioned positively or negatively developed a more recent visualization system called Emotion Watch that constructs visual sum-maries of public emotions. They applied it to the 2012 Olympics as a test case and have shown that users prefer a more detailed inspection of public emotions over the simplified analysis. Besides these automated tools, various online tools like Twitrratr, Twendz, Social mention, and Sentimetrics are avail-able to track the sentiment in social networks. The competition among topics on social media are modeled and visualized in the visual analysis system produced by [86]. [13] utilizes an improved stacked graph to better understand large-scale events on Twitter and allow users to visualize the topics extracted from streaming tweets. The Topic Streams support three basic filter interactions: temporal zoom, temporal sliding and topical selection. All these inter-actions result in animated transitions of the Topic Streams and the other views, reducing the cognitive cost of following the change. We remark that the above approaches focus on visual exploration of evolving topics, including content change and strength change. Unlike these approaches, [48] propose a system that illustrate opinion diffusion pattern on social media. An opinion flow visualization that combines a Sankey diagram with a density map is presented to help users better understand the complex opinion diffusion process. The user migration between subtopics of the same higher level topic is the unique feature visualized by this system. It would be straightforward to extend it to visualize user migration between subtopics of different higher level topics. However, the benefits and use scenarios of this feature are still unclear. Furthermore, this feature could lead to visual clutter because there would be too many topic strips and transition. Opinion flow supports a rich set of user interactions for the in-depth analysis of opinion propagation to meet design goal G3: Select Users for Detailed Examination, Interact with The Visual Topic Tree, Trace Influence of Users on Opinion Diffusion, Navigate A Multi-scale Timeline, Examine The Diffusion Behavior of Users, Validate The "What If" Hypothesis, Remove Smaller Topics and Transitions. [87] has created a prototype 2 which is designed for evaluating customer sentiment regarding hotels. The underlying techniques used are fairly general, and could be applied to any relevant and sufficiently large data set containing sentiment data. To fully utilize the prototype, the only requirement is that the data set contains both a geographical and a temporal aspect. However, all techniques and features can also be used separately. Feature search and extraction may be used on any opinionated data set, although presentation of results in the prototype would need altering to some sort of ranked list, for instance. Burst detection may be used on any temporal data set, and might even function better on other types of data sets, for instance movie or video game reviews.

V. Evaluation & Discussion

We have presented a survey of visual sentiment analytics systems to help users better understand the evolution of the visualization field at different levels of granularity. Visualization of opinions has not yet been a major focus of research in the area. Despite the increase of interest in sentiment analysis, many tools do not pay much attention to the user interface aspects. These aspects are very important in order to satisfy the user needs. To the best of our knowledge currently no system exists that analyzes and visualizes customer feedback with respect to clusters of reviews in which similar opinions are expressed. Existing systems have achieved certain success, but they cannot enable analysts to make a quick decision. In addition, these tools provide scant support for complex opinion analysis such us identifying factors influencing customer behaviors. This is a huge advancement in brands' ability to understand why social trends and spikes are happening so companies can act on it to create better content and stronger connections with their customers (the Why). In addition, most existing sentiment visualization fails to meet all the requirements simultaneously. We remark from the above study that only few summarization systems put the analysis and summary of people's emotional reactions at the center of their work; rather they presented sentiments as additional dimensions of content. Moreover, all systems showed sentiment in terms of polarity classifications only. To address these challenges, a fine-grained visual representation of emotions should be designed. Each system is characterized by its applicability, advantages, and disadvantages. Most other systems focus on reconstructing the event's structure and sub-events, and not focusing on presenting people's emotional reactions to the event. Further more, they summarize the sentiment polarity of tweets on just polarity leaving a more detailed and insightful analysis to be desired. We suggest using a multi-category summary of emotional reactions using a fine-grained set of categories because users found it beneficial to distinguish be-tween separate types of emotions instead of just polarity. The study also showed a problem in the visual representation of the interface itself.

VI. CONCLUSION

We have presented a concise review and a state-of-the art report presenting recent approaches in the field of sentiment visualization. In particular, we were focused on witnessing rapid advances in opinion visualization techniques and tools to gain a better understanding of the field. From this review, al-though these approaches have proven their usefulness in opinion visualization field, they suffer from some drawbacks. There are likely to be many other applications that are not discussed. It is found that sentiment visualization tools and techniques are severely de-pendent on topics and objectives. From the above survey it is evident that neither technique model consistently outperforms the other, different types of methods have distinct objectives. It is also found that different types of visualization techniques are combined in an efficient way in order to overcome their individual drawbacks and benefit from each other's merits, and finally enhance the sentiment visualization performance. In future, more work is needed on further improving the performance measures. Sentiment analysis can be applied for new applications. Although the techniques and algorithms used for sentiment visualization are advancing fast, however, a lot of problems in this field of study remain unsolved. More future research could be dedicated to these challenges.

REFERENCES

[1] L. A. Adamic and E. Adar, "Friends and neighbors on the web," *Social Networks 25*, no. 3, pp. 211–230, 2003.

[2] A. Bifet and E. Frank, "Sentiment knowledge discovery in twitter streaming data," in *Proc. the 13th International Conference on Discovery Science*, Berlin, Heidelberg, Springer-Verlag, 2010, pp. 1–15.

[3] E. Bjørkelund, T. H. Burnett, and K. Nørvåg, "A study of opinion mining and visualization of hotel reviews," in *Proc. 14th International Conference on Information Integration and Web-based Applications & Services*, Bali, Indonesia, December 3-5, 2012, pp. 229–238.

[4] C. Blundell, Y. W. Teh, and K. A. Heller, Bayesian Rose Trees, CoRR abs/1203.3468 (2012).

[5] N. Cao, Y. R. Lin, X. H. Sun, D. Lazer, S. X. Liu, and H. M. Qu, "Whisper: Tracing the spatiotemporal process of information diffusion in real time," *IEEE Trans. Vis. Comput. Graph*, vol. 18, no. 12, pp. 2649–2658, 2012.

[6] C. Chelmis and V. K. Prasanna, "Social networking analysis: A state of the art and the effect of semantics," *IEEE on Social Com/PASSAT*, 2011, pp. 531–536.

[7] C. Chen, F. Ibekwe-Sanjuan, E. S. Juan, and C. Weaver, "Visual analysis of conflicting opinions," in *Proc. IEEE VAST'06*, 2006, pp. 59–66.

[8] W. W. Cui, S. X. Liu, L. Tan, C. L. Shi, Y. Q. Song, Z. K. Gao, H. M. Qu, and X. Tong, "Textflow: Towards better understanding of evolving topics in text," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, 2412–2421, 2011.

[9] W. W. Cui, S. X. Liu, Z. F. Wu, and H. Wei, "How hierarchical topics evolve in large text corpora," *IEEE Trans. Vis. Comput. Graph*, vol. 20, no. 12, pp. 2281–2290, 2014.

[10] W. W. Cui, Y. C. Wu, S. X. Liu, F. Wei, M. X. Zhou, and H. M. Qu, "Context preserving, dynamic word cloud visualization," *IEEE Computer Graphics and Applications* vol. 30, no. 6, pp. 42–53, 2010.

[11] L. Ding, L. Zhou, T. Finin, and A. Joshi, "How the semantic web is being used: An analysis of foaf documents," in *Proc. 38th Annual Hawaii International Conference on System Sciences (HICSS'05)-Track 4, IEEE Computer Society*, vol. 04 Washington, DC, USA, 2005, pp. 113.3.

[12] X. W. Ding and B. Liu, "The utility of linguistic rules in opinion mining," in *Proc. the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA, pp. 811–812.

[13] X. W. Ding, B. Liu, and P. S. Yu, *A Holistic Lexicon-based Approach to Opinion Mining*, 2008.

[14] M. Dork, D. Gruen, C. Williamson, and S. Carpendale, "A visual backchannel for large-scale events," IEEE *Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1129–1138, 2010.

[15] W. W. Dou, L. Yu, X. Y. Wang, Z. Q. Ma, and W. Ribarsky, "Hierarchical topics: Visually exploring large text collections using topic hierarchies," *IEEE Trans. Vis. Comput. Graph*, pp. 2002–2011, 2013.

[16] G. M. Draper and R. F. Riesenfeld, "Who votes for what? A visual query language for opinion data," *IEEE Trans. Vis. Comput. Graph*, vol. 14, no. 6, 1197–1204, 2008.

[17] G. Erétéo, M. Buffa, F. Gandon, P. Grohan, M. Leitzelman, and P. Sander, "A state of the art on social network analysis and its applications on a semantic web," 2008.

[18] S. Faridani, E. Bitton, K. Ryokai, and K. Goldberg, "Opinion space: A scalable tool for browsing online comments," in *Proc. the SIGCHI Conference on Human Factors in Computing Systems*, New York, USA, 2010, pp. 1175–1184.

[19] D. Fisher, A. Hoff, G. G. Robertson, and M. Hurst, "Narratives: A visualization to track narrative events as they develop," *IEEE VAST, Computer Society*, pp. 115–122, 2008.

[20] T. Fukuhara and T. Nishida, "Understanding sentiment of people from news articles: Temporal sentiment analysis of social events," in *Proc. ICWSM*, 2007.

[21] M. Gamon, A. Aue, S. C. Oliver, and E. Ringger, "Pulse: Mining customer opinions from free text," in *Proc. the 6th International Conference on Advances in Intelligent Data Analysis*, Berlin, Heidelberg), Springer-Verlag, 2005, pp. 121–132.

[22] M. L. Gregory, N. Chinchor, P. Whitney, R. Carter, E. Hetzler, and A. Turner, "User-directed sentiment analysis: Visualizing the affective content of documents," in *Proc. the Workshop on Sentiment and Subjectivity in Text*, Sydney, Australia, Association for Computational Linguistics, July 2006, pp. 23–30.

[23] "User-directed sentiment analysis: Visualizing the affective content of documents," in *Proc. the Workshop on Sentiment and Subjectivity in Text*, Sydney, Australia, Association for Computational Linguistics, July 2006, pp. 23–30.

[24] M. Hao, U. Dayal, D. Keim, and T. Schreck, "Multi-resolution techniques for visual exploration of large time-series data," in *Proc. Eurographics/ IEEE-VGTC Symposium on Visualization*, K. Museth, T. Moeller, and A. Ynnerman, eds. The Eurographics Association, 2007.

[25] M. C. Hao, C. Rohrdantz, H. Janetzko, D. A. Keim, U. Dayal, L. Haug, M. Hsu, and F. Stoffel, "Visual sentiment analysis of customer feedback streams using geo-temporal term associations," *Information Visualization*, vol. 12, no. 3-4, pp. 273–290, 2013.

[26] S. Havre, E. Hetzler, P. Whitney, and L. Nowell, "Themeriver: Visualizing thematic changes in large document collections," *IEEE Transactions on Visualization and Computer Graphics*, vol. 8, no. 1, pp. 9–20, 2002.

[27] M. Hu, H. H. Yang, M. X. Zhou, L. Gou, Y. Y. Li, and E. M. Haber, "Opinionblocks: A crowd-powered, selfimproving interactive visual analytic system for understanding opinion text," in *Proc. 14th IFIP TC 13 International Conference Human-Computer Interaction-Interact*, Cape Town, South Africa, September 2-6, 2013, pp. 116–134.

[28] A. Kaur and V. Gupta, "A survey on sentiment analysis and opinion mining techniques," *Journal of Emerging Technologies in Web Intelligence*, vol. 5, no. 4, 2013.

[29] H. A. Kautz, B. Selman, and M. A. Shah, "The hidden web, AI Magazine," vol. 18, no. 2, pp. 27–36, 1997.

[30] R. Kempter, V. Sintsova, C. Musat, and P. Pu, "Emotion Watch: Visualizing fine-grained emotions in event-related tweets," 2014.

[31] K. Koh, B. Lee, Bo H. Kim, and J. Seo, "Maniwordle: Providing flexible control over wordle," *IEEE Trans. Vis. Comput. Graph*., vol. 16, no. 6, pp. 1190–1197, 2010.

[32] J. Leskovec, L. Backstrom, and J. Kleinberg, "Meme-tracking and the dynamics of the news cycle," in *Proc. the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, USA, 2009, pp. 497– 506.

[33] W. Li, L. X. Duan, D. Xu, and I. W. Tsang, "Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 6, pp. 1134–1148, 2014.

[34] C. Lin, Y. He, R. Everson, and S. Ruger, "Weakly supervised joint sentiment-topic detection from text," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 6, pp. 1134–1145, 2012.

[35] B. Liu, "Sentiment analysis and subjectivity," *Handbook of Natural Language Processing*, 2nd Ed. Taylor and Francis Group, Boca, 2010.

[36] B. Liu, M. Q. Hu, and J. S. Cheng, "Opinion observer: Analyzing and comparing opinions on the web," in *Proc. the 14th International Conference on World Wide Web*, New York, USA, 2005, pp. 342–351.

[37] S. X. Liu, W. W. Cui, Y. C. Wu, and M. C. Liu, "A survey on information visualization: Recent advances and challenges," *Vis. Computer*, vol. 30, no. 12, pp. 1373–1393, 2014.

[38] S. X. Liu, X. T. Wang, J. F. Chen, J. Zhu, and B. N. Guo, "Topic panorama: A full picture of relevant topics," in *Proc. 2014 IEEE Conference on Visual Analytics Science and Technology*, Paris, France, October 25-31, 2014, pp. 183–192.

[39] S. X. Liu, M. X. Zhou, S. M. Pan, W. H. Qian, W. J. Cai, and X. X. Lian, "Interactive, topic-based visual text summarization and analysis," in *Proc. the 18th ACM Conference on Information and Knowledge Management*, New York, USA, 2009, pp. 543–552.

[40] S. X. Liu, M. X. Zhou, S. M. Pan, Y. Q. Song, W. H. Qian, W. J. Cai, and X. X. Lian, "Tiara: Interactive, topic-based visual text summarization and analysis," *ACM Trans. Intell. Syst. Technol*., vol. 3, no. 2, pp. 25:1–25:28, 2012.

[41] D. N. Luo, J. Yang, M. Krstajic, W. Ribarsky, and D. A. Keim, "Eventriver: Visually exploring text collections with temporal references," *IEEE Trans. Vis. Comput. Graph*., vol. 18, no. 1, pp. 93-105, 2012.

[42] A. Marcus, M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, and R. C. Miller, "Twitinfo: Aggregating and visualizing microblogs for event exploration," in *Proc. the SIGCHI Conference on Human Factors in Computing Systems*, New York, USA, 2011, pp. 227-236.

[43] Q. L. Miao, Q. D. Li, and R. W. Dai, "Amazing: A sentiment mining and retrieval system," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 7192–7198, 2009.

[44] P. Mika, "Social networks and the semantic web," *Semantic Web and Beyond*, vol. 5, Springer, Berlin, 2007.

[45] S. M. Kim and E. Hovy, "Determining the sentiment of opinions," in *Proc. the 20th International Conference on Computational Linguistics*, 2004, pp. 1367–1373.

[46] G. Mishne and M. D. Rijke, "Moodviews: Tools for blog mood analysis," in *Proc. AAAI Symposium on Computational Approaches to Analysing Weblogs*, 2006, pp. 153–154.

[47] G. Mishne and M. D. Rijke, "Moodviews: Tools for blog mood analysis," in *Proc. AAAI 2006 Spring Symposium on Computational Approaches*, 2006.

[48] M. Monroe, R. J. Lan, H. Lee, C. Plaisant, and B. Shneiderman, "Temporal event sequence simplification," *IEEE Trans. Vis. Comput. Graph*., vol. 19, no. 12, pp. 2227–2236, 2013.

[49] S. Morinaga, K. Yamanishi, K. Tateishi, and T. Fukushima, "Mining product reputations on the web," in *Proc. the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, USA, 2002, pp. 341–349.

[50] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, "Sentiful: A lexicon for sentiment analysis," *T. Affective Computing*, vol. 2, no. 1, pp. 22–36, 2011.

[51] M. Newman, *Networks: An Introduction*, Oxford University Press, Inc., New York, NY, USA, 2010.

[52] V. Ng, S. Dasgupta, and S. M. N. Arifin, "Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews," in *Proc. COLING/ ACL 2006 Main Conference Poster Sessions*, 2006, pp. 611–618.

[53] D. Oelke, M. C. Hao, C. Rohrdantz, D. A. Keim, U. Dayal, L. E. Haug, and H. Janetzko, "Visual opinion analysis of customer feedback data," in *Proc. the IEEE Symposium on Visual Analytics Science and Technology*, Atlantic City, New Jersey, USA, October 11-16, 2009, pp. 187–194.

[54] D. Oelke, M. C. Hao, C. Rohrdantz, D. A. Keim, U. Dayal, L.E. Haug, and H. Janetzko, "Visual opinion analysis of customer feedback data," *IEEE VAST'09*, 2009, pp. 187–194.

[55] D. Oelke, D. Spretke, A. Stoffel, and D. A. Keim, "Visual readability analysis: How to make your writings easier to read," *IEEE Trans. Vis. Comput. Graph*, vol. 18, no. 5, pp. 662–674, 2012.

[56] M. Ogawa and K. L. Ma, *Software Evolution Storylines*, SOFTVIS (Alexandru Telea, Carsten Görg, and Steven P. Reiss, eds.), ACM, 2010, pp. 35–42.

[57] T. H. Ong, H. Chen, W. Sung, and B. Zhu, "Newsmap: A knowledge map for online news," *Decision Support Systems*, vol. 39, no. 4, pp. 583–597, 2005.

[58] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr*., vol. 2, no. 1-2, pp. 1–135, 2008.

[59] A. M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in *Proc. the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA, HLT '05, Association for Computational Linguistics, 2005, pp. 339–346.

[60] G. Qiu, B. Liu, J. J. Bu, and C. Chen, "Expanding domain sentiment lexicon through double propagation," in *Proc. the 21st International Jont Conference on Artifical Intelligence*, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc., 2009, pp. 1199–1204.

[61] E. Riloff and J. Wiebe, "Learning extraction patterns for subjective expressions," in *Proc. the 2003 Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA, Association for Computational Linguistics, 2003, pp. 105–112.

[62] M. Rios and J. Lin, "Distilling massive amounts of data into simple visualizations: Twitter case studies," 2012.

[63] C. Rohrdantz, M. C. Hao, U. Dayal, L. E. Haug, and D. A. Keim, "Feature-based visual sentiment analysis of text document streams," *ACM Trans. Intell. Syst. Technol*. vol. 3, no. 2, pp. 26:1–26:25. 2012.

[64] "Feature-based visual sentiment analysis of text document streams," *ACM Trans. Intell. Syst. Technol.* vol. 3, no. 2, pp. 26:1–26:25, 2012.

[65] S. J. Rose, S. Butner, W. Cowley, M. L. Gregory, and J. Walker, "Describing story evolution from dynamic information streams," *IEEE VAST, IEEE Computer Society*, pp. 99–106, 2009.
[66] M. A. Smith and A. T. Fiore, "Visualization components for persistent conversations," in *Proc. the SIGCHI Conference on Human Factors in Computing Systems*, New York, USA, CHI '01, ACM, 2001, pp. 136–143.
[67] H. Strobelt, D. Oelke, C. Rohrdantz, A. Stoffel, D. A. Keim, and O. Deussen, "Document cards: A top trumps visualization for documents," *IEEE Trans. Vis. Comput. Graph*. vol. 15, no. 6, pp. 1145–1152, 2009.
[68] K. Sznajd-Weron, M. Tabiszewski, and A. M. Timpanaro, "Phase transition in the sznajd model with independence," CoRR abs/1106.0934, 2011.
[69] Y. Tanahashi and K. Ma, "Design considerations for optimizing storyline visualizations," *IEEE Trans. Vis. Comput. Graph*., vol. 18, no. 12, pp. 2679–2688, 2012.
[70] "Design considerations for optimizing storyline visualizations," *IEEE Trans. Vis. Comput. Graph*. vol. 18, no. 12, pp. 2679–2688, 2012.
[71] J. R. Tyler, D. M. Wilkinson, and B. A. Huberman, "Email as spectroscopy: Automated discovery of community structure within organizations," pp. 81–96, 2003.
[72] F. B. Viegas, M. Wattenberg, and J. Feinberg, "Participatory visualization with wordle," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1137–1144, 2009.
[73] T. V. Landesberger, A. Kuijper, T. Schreck, J. Kohlhammer, J. J. V. Wijk, J. Fekete, and D. W. Fellner, "Visual analysis of large graphs: State-ofthe- art and future research challenges," *Comput. Graph. Forum*, vol. 30, no. 6, pp. 1719–1749, 2011.
[74] C. Wang, Z. Xiao, Y. H. Liu, Y. R. Xu, A. Y. Zhou, and K. Zhang, "Sentiview: Sentiment analysis and visualization for internet popular topics," *IEEE T. Human-Machine Systems*, vol. 43, no. 6, pp. 620–630, 2013.
[75] F. Wanner, C. Rohrdantz, F. Mansmann, D. Oelke, and D. A. Keim, "Visual sentiment analysis of rss news feeds featuring the us presidential election in 2008," in *Proc. Workshop on Visual Interfaces to the Social and the Semantic Web*, 2009.
[76] S. Wasserman and K. Faust, "Social network analysis: Methods and applications," vol. 8, Cambridge University Press, 1994.
[77] K. Wongsuphasawat and D. Gotz, "Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization," *IEEE Trans. Vis. Comput. Graph*, vol. 18, no. 12, pp. 2659–2668, 2012.
[78] K. Wongsuphasawat, J. A. G. Gómez, C. Plaisant, T. D. Wang, M. Taieb-Maimon, and B. Shneiderman, "Lifeflow: Visualizing an overview of event sequences," in *Proc. the SIGCHI Conference on Human Factors in Computing Systems*, New York, USA, CHI '11, ACM, 2011, pp. 1747–1756.
[79] Y. C. Wu, S. X. Liu, K. Yan, M. C. Liu, and F. Z. Wu, "Opinionflow: Visual analysis of opinion diffusion on social media," *IEEE Trans. Vis. Comput. Graph*, vol. 20, no. 12, pp. 1763–1772, 2014.
[80] Y. C. Wu, T. Provan, F. Wei, S. X. Liu, and K. Ma, "Semantic-preserving word clouds by seam carving," *Computer Graphics Forum*, vol. 30, no. 3, pp. 741–750, 2011.
[81] Y. C. Wu, F. Wei, S. X. Liu, N. Au, W. W. Cui, H. Zhou, and H. Qu, "Opinionseer: Interactive visualization of hotel customer feedback," *IEEE Trans. Vis. Comput. Graph*, vol. 16, no. 6, pp. 1109–1118, 2010.
[82] K. Q. Xu, S. Shao, Y. Liao, J. X. Li, and Y. X. Song, "Mining comparative opinions from customer reviews for competitive intelligence," *Decision Support Systems*, vol. 50, no. 4, pp. 743–754, 2011.
[83] P. P. Xu, Y. C. Wu, E. X. Wei, T. Peng, S. X. Liu, J. J. H. Zhu, and H. Qu, "Visual analysis of topic competition on social media," *IEEE Trans. Vis. Comput. Graph*, vol. 19, no. 12, pp. 2012–2021, 2013.
[84] "Visual analysis of topic competition on social media," *IEEE Trans. Vis. Comput. Graph*, vol. 19, no. 12, pp. 2012–2021, 2013.
[85] Z. J. Yin, M. Gupta, T. Weninger, and J. W. Han, "A unified framework for link recommendation using random walks," ASONAM Nasrullah Memon and Reda Alhajj, eds., in *IEEE Computer Society*, 2010, pp. 152–159.
[86] C. H. Zhang, Y. H. Liu, and C. B. Wang, "Time-space varying visual analysis of micro-blog sentiment," in *Proc. VINCI'13*, 2013, pp. 64–71.
[87] Z. Q. Zhang, Q. Ye, Z. L. Zhang, and Y. J. Li, "Sentiment classification of internet restaurant reviews written in cantonese," *Expert Syst. Appl*., vol. 38, no. 6, pp. 7674–7682, 2011.