CommentVis: Unveiling Comment Insights Through Interactive Visualization Tool

Guangjing Yan* [1], Jinhwa Jang* [1 2†], and Jinwook Seo‡ [1]

[1]Seoul National University [2]Samsung Electronics

Figure 1: CommentVis consists of five linked components: (A) *Comparative Dashboard* exhibits (A1) dual scatterplot snapshots alongside their metadata and (A2) *Comparison Summary* produced by GPT-4-Turbo, highlighting contextual shifts over time. (B) *Sentiment Trend Timeline* provides an intuitive representation of how sentiment fluctuates over time. (C) *Overview Scatterplot* shows the semantic distribution of comment embedding, augmented by interactive exploration tools such as (C1) Tooltip and (C2) Brush. (D) *Topics Navigator* displays hot topics extracted from *selected comments*. (E) *Abstract Generator* presents (E1) the abstract of *selected comments*, generated by GPT-4-Turbo, and lists (E2) Top Comments. When the mouse hovers over the listed comments, the corresponding points in *Overview Scatterplot* will be marked with a red border. *Selected comments* refer to the results of user interactions through the (B) *Sentiment Trend Timeline*, (C2) *Brush tool*, or (D) *Topics Navigator* for a refined analysis.

Abstract

Recently, online forums have emerged as a stage for consumers to comment and share their reviews. These user comments serve as a valuable data source for marketing professionals and analysts. Nevertheless, conventional user interfaces often present an overwhelming volume of comments in a linear-structured list, significantly impeding the efficiency of marketing professionals in analyzing the feedback. In response to this challenge, we introduce CommentVis, an interactive visualization tool that helps users grasp the skeleton of comments' semantic distribution. Using the tool, analysts gain detailed information about comments with profound insights. The tool leverages state-of-the-art large-scale language models, optimizing the speed and depth of analysis of substantial text data that may take a long time for marketers. To illustrate the practical application of CommentVis, we present a usage scenario that demonstrates its effectiveness in real-world marketing analysis. The tool's impact and utility were further validated through a user study involving three marketing professionals in a global manufacturing company.

Index Terms:
Human-centered computing—Visualization—Visualization systems and tools—; Computing methodologies—Artificial intelligence—Natural language processing—;

*Guangjing Yan and Jinhwa Jang equally contributed to this work. e-mail: {gjyan, jhjang}@hcil.snu.ac.kr

†jinfire.jang@samsung.com

‡jseo@snu.ac.kr, corresponding author

1 Introduction

Online comments discussing specific products on social media such as Reddit, X, and Facebook are often considered to authentically reflect consumers' preferences and dislikes toward the products. They reveal the product's strengths and weaknesses after undergoing market scrutiny and play a significant role in influencing the purchase patterns of consumers [8]. Therefore, analyzing a large number of online comments has become a crucial method for marketers to grasp public attitudes toward their products. Marketers extract useful information from comments based on attributions like sentiments, keywords, received likes, and posting times for market analysis.

However, analysis of comments data is demanding as analysts are required to examine a substantial amount of data presented in a linear, sequential format [20]. This conventional display lacks a comprehensive overview, sorting only by specific criteria such as posting time and the number of likes received, thereby making the analytical process laborious and time-consuming. In this way, marketers are unable to quickly discern the overall skeleton of contents, making it challenging to capture insights. To alleviate these limitations, prior works [4, 6] were proposed. Still, either of them suffers from a lack of the capability to delve deeper and extract more granular information or is difficult to comprehend due to a lack of labels and contextual guidance.

To fill this research gap, we propose a visualization tool, CommentVis (Fig. 1), supporting users to grasp the high-level synthesis of comment contents while being able to extract detailed information of insightful comments. Our tool employs visualization techniques

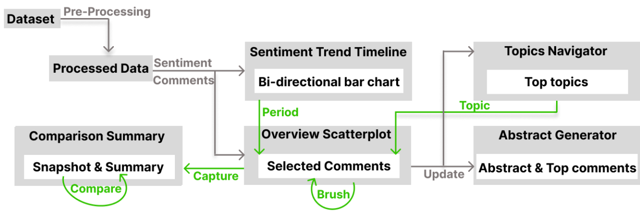

Figure 2: The workflow diagram of CommentVis illustrates the data flow, indicated by grey arrows, and user interactions, denoted by green arrows. The views display initial processed data by default. Selected comments filtered in *Overview Scatterplot* by user interactions trigger the update of *Topics Navigator* and *Abstract Generator*.

such as interactive scatterplot, dimension reduction, and state-of-the-art Large Language Models (LLM) [18] to accelerate users' exploration and acquisition of insight from a mass of comments. Workflow and relationships between views are illustrated in Fig. 2.

The development of CommentVis was informed by a preliminary study with marketing professionals at a global manufacturing firm. This collaboration was instrumental in understanding the standard analysis practices and user needs in market analysis. The insights and feedback obtained from these professionals played a crucial role in shaping the design and functionality of CommentVis, ensuring it effectively supports users in synthesizing high-level content from comments and extracting detailed insights. The integration of visualization and LLMs was directly influenced by the requirements and challenges identified through this preliminary study.

The efficacy of CommentVis was assessed through qualitative studies with experienced marketing professionals using task-based scenarios. The results validate that CommentVis enhances analytical efficiency by providing intuitive visual interactions coupled with large language models, providing a comprehensive overview, spotlighting salient discussions, and condensing large text volumes into summaries. Post-hoc interviews revealed that even though the professionals have limited expertise in visualization, they were able to readily handle dense comments data using CommentVis.

In conclusion, our contributions are as follows:

- We propose a workflow that combines interactive visual analysis with a large-scale language model to extract insightful features from text embeddings.
- We implement CommentVis as an interactive visualization tool that helps users effectively explore text embeddings while also maintaining enough trending context to help users identify critical events.
- We conduct qualitative user studies with marketing professionals, confirming that CommentVis improves performance in analyzing a large volume of comments through intuitive visualization and is well-received for its ability to manage dense data and ease of use.

2 Related Work

Text Embedding Generation. A text embedding is a projection of text in high-dimensional latent space. Similar texts are supposed to be positioned closer than less similar texts in the embedding space. With the fast development of natural language processing techniques, more and more advanced text embedding models were proposed, and they have been confirmed to be able to preserve the sentence's semantic meaning well. Among them, BERT [5] achieves a new state-of-the-art performance in sentence embedding.

In our work, text-embedding-ada-002 from OpenAI is used to generate comment embeddings utilized in constructing a scatterplot. This choice is not only because it's the encoder of GPT-4-turbo [18], which is used in our system to extract the abstract of comments, but also because it outperforms the aforementioned BERT-based embedding model and achieves high performance in various benchmarks with its ability to embed up to 6000 words [17].

Sentiment Analysis. Sentiment analysis has employed techniques from simple lexicon-based methods to complex machine learning algorithms, including Naive Bayes and SVM. Recently, deep learning approaches like LSTM and CNN have made significant strides. However, the transformer architecture of RoBERTa models [14], trained on Twitter datasets, has proven superior in understanding the context and informality typical of social media like Reddit. This aligns with our analysis objectives, rendering RoBERTa an optimal choice for assessing nuanced sentiment in user-generated content.

Visual Analytics for Comments. Using scatterplot to display 2D nonlinear projection of text-embedding is a conventional way to visualize text structure and extract clusters and other in-depth insights. Opinion Space proposed by Faridani et al. [6] combined 2D scatterplot view with collaborative filtering, has confirmed that displaying comments in this way highlights the most insightful comments and freely exploring it enables the user to see more diversity among comments in a user study comparing to grid and list format. However, people also want to explore comments from other perspectives, for example, time. To grasp the flow of topic evolution, prior work made by Weiwei Cui et al.has proposed an interactive visualization tool called TextFlow [4]. Combining a flow graph with 4 specific glyphs and keyword thread, it allows users to track and connect topics over time, identify critical events and keyword correlations, and visually convey complex relationships between them.

While these useful tools have been proposed, none of them provides overall topic and sentiment evolution while being able to extract detailed information. Our work aims to address this issue by introducing CommentVis, which combines a scatterplot that presents the structure of the text in the semantic space with a timeline that indicates fluctuation of sentiment trend. By observing the clusters and distributions within the scatterplot, we gain instructions for the next steps in our in-depth exploration.

3 Domain Situation and Design Requirements

To identify key challenges in processing online consumer feedback, we engaged with two marketing professionals at a global manufacturing company with 8 and 9 years of experience, respectively. They provided the traditional method of analyzing text comments from platforms like Reddit and Facebook. This time-consuming and inefficient practice involves a verbose and poorly categorized linear structure. The process typically starts with marketers selecting related posts, reading through a large volume of comments, and sampling while manually filtering out irrelevant content.

As a result, we found that the filtered comments form the basis for downstream analysis. These comments are categorized based on specific times and topics and then organized into lists for further examination to uncover underlying patterns. However, this traditional approach is often inefficient and demands substantial expertise from analysts. The inefficiency stems from the absence of tools for data filtering and analysis, leading marketing personnel to rely heavily on subjective judgment throughout the entire process.

To address these challenges, we concentrated on developing a solution that offers (1) a comprehensive explanation of overall review comments and (2) intuitive interaction designs for exploring specific features based on sentiments, topics, and periods. This led to the establishment of four core design requirements for CommentVis:

- **DR1: Offer a comprehensive overview of synthesis information.** Target users struggle to efficiently discern valuable topics and comments with profound insight from a large volume of comments. Thus, the system should provide an overview that assists users in efficiently grasping high-level synthesis of comment information and providing guidance to extract meaningful patterns from a given context.
- **DR2: Automatically distill lengthy information and provide concise summaries.** Users seek an automated tool to help analyze a substantial volume of comments. Their traditional

manual approach not only demands excessive effort but also heavily relies on the marketers' experience, leading to potential errors or omissions. Therefore, the system should offer an automated tool to efficiently assist users in making objective discoveries through interaction.

- **DR3: Dig into mainstream topics and provide means of tracing back.** For users, assessing and summing up user feedback involves deducing mainstream opinions from a large number of comments, and this can be time-consuming. To expedite this process, our system should offer automated categorization of mainstream opinions and provide means for users to trace the origins of perspectives. This facilitates users in identifying the underlying reasons for a specific viewpoint becoming mainstream.
- **DR4: Present the trend of public sentiment fluctuating over time.** Monitoring public sentiment before and after the launch of a new product is crucial for marketers to promptly understand which features of the product are more appealing to consumers and what event sparked intense public discussion. This information can serve as a focal point for marketing efforts and provide insights into any shortcomings in a particular aspect of the product, allowing for optimization in subsequent product releases. Therefore, the system should present a sentiment change over time that allows the users to identify crucial events and topics that arouse sentiment fluctuation.

These design requirements are crafted to transform extensive consumer feedback into actionable insights, thereby equipping marketing professionals with an advanced, user-friendly analytical tool. They reflect a deep understanding of the practical challenges and needs in marketing analytics, ensuring that CommentVis effectively addresses the demands of modern marketing environments.

4 Data Preprocessing and Workflow

Our preprocessing pipeline for CommentVis integrates advanced methodologies to meet the complexities of analyzing online comments that mainly consist of consumer feedback and extracting meaningful information from them.

Data Crawling and Automatic Filtering. We collected raw comment data using PRAW [1], a Python Reddit API wrapper, which collects the main information of the comment, including comment text, posted time, posted user, and received likes. However, due to the complexity of comments in the online forum, the raw data crawled from Reddit often contains a significant amount of tangential discussions unrelated to the main topic. These noises can have a bad impact when observing text embedding clustering in subsequent downstream tasks, but manually removing these irrelevant comments can be laborious and time-consuming. Therefore, we utilized GPT-4-Turbo [18] to filter the comment data by combining the post's title and content as prompts listed in Table 2.

Text embedding and Dimensionality Reduction. We utilized text embeddings generated from OpenAI's language model, text-embedding-ada-002 [2] to capture nuanced text semantics, which builds the foundation for subsequent tasks. The 2D position of text embedding that could be used in the following visualization task for disclosing cluster patterns (DR1) is accessed through the nonlinear dimensionality reduction technique UMAP [15]. We used UMAP as it has been widely validated well to preserve the cluster structure of the original high-dimensional data [11]. However, in order to minimize the occurrence of distortion, we utilize ZADU [10] to assess the reliability of the embeddings through Trustworthiness & Continuity [13], and fine-tuning parameters in UMAP according to the score of distortion measures with the help of Bayesian Optimization[2]. We did not utilize clustering algorithms like k-means to provide beforehand prediction that classifies data into a specific number of clusters. Instead, we chose to plainly encode 2D projection of comment embedding to the position of dots to respect the complexity and multifaceted nature of consumer feedback, recognizing consumer comments often encompass multiple overlapping topics. This decision aligns with the challenges in text clustering [16].

[1]https://github.com/praw-dev/praw

[2]https://github.com/bayesian-optimization/BayesianOptimization

Functionality	Initial Latency	Cache Key	Cache Latency
Generate Abstract	25.6s	Period *p1*; Topic *t1*;	0.21s
Comparison Summary	27.2s	Periods *p1*, *p2*; Topics *t1*, *t2*;	0.22s

Table 1: Caching System Specifications. Latency measurements were derived from 50 individual requests, quantified by Python scripting.

Purpose	Prompt
Automatic Filtering	Identify if *comments* correspond to their respective *posts*, and indicate "related" or "not related".
Generate Abstract	Summarize *product comments* in three sentences, focusing on sentiment and key features, with highlights for marketers
Comparison Summary	Compare *comments* from *periods p1 and p2* or *topics t1 and t2* focusing on shifts in sentiment, highlighting key changes.

Table 2: Concise Variations of GPT-4-Turbo Prompts.

Topic Extraction. We used the KeyBERT [7] to extract topics from customer review comments. The KeyBERT is designed to extract relevant topics efficiently. To identify common topics, we aggregated comments and used the Maximal Marginal Relevance [1] method to reduce the diversity of expressions to the same topic. We then extracted ten topics from each comment and filtered those to common topics (DR3). This was done because social media text often contains several topics [9]. This method not only prevents comments from being narrowly categorized under a single topic but also facilitates successive topic-based data filtering without omitting relevant comments. Additionally, this pre-processing enables the extraction of trend topics of given conditions, such as periods.

Sentiment analysis. We analyzed the sentiment of the extracted comments using the state-of-art sentiment analysis pre-training model [14], classifying each comment sentiment into three categories, positive, negative and neutral. The result of sentiment classification can be grouped if topics or periods are given; it is used to display the distribution of sentiment per topic or period (DR4).

Generate Summary. Summarizing extensive comments from online platforms is challenging. As a result, LLMs have become widely used for generating abstracts from text data [3] (DR2). We chose the state-of-art model GPT-4-Turbo [18] for this task due to its large text window capacity of up to 128K, enabling effective summarization of vast amounts of comments. During iterative design sessions with marketing professionals, latency in summary generation was identified as a key issue impacting user experience. To address it, the caching system was devised. It ensures that texts generated on a first request are cached, enabling significantly faster responses for subsequent requests. Details of the system are provided in Table 1. In addition, to enhance content engagement, we tailored prompts used in CommentVis for different purposes, detailed in Table 2.

By integrating these methodologies, our preprocessing pipeline ensures that CommentVis captures and accurately represents the dynamic nature of online consumer feedback, readying the data for effective visualization and analysis.

5 Visualization Design

Based on the design requirements proposed in Section 3, we developed a visualization system called CommentVis, aiming to accelerate the process of marketers analyzing the market using online comments. Our system consists of the following five components:

Sentiment Trend Timeline. Sentiment Trend Timeline (Fig. 1.B)

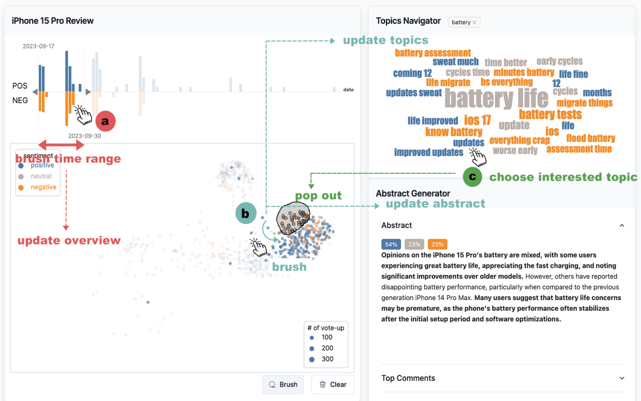

Figure 3: Linking between *Sentiment Trend Timeline*, *Overview Scatterplot*, *Topics Navigator*, and *Abstract Generator*. (a) Users can brush the timeline to see comments posted during the selected time range in the scatterplot. (b) The mainstream topics and distilled abstract of comments brushed by users using the lasso tool will be shown in *Topics Navigator* and *Abstract Generator*, respectively. (c) The comments related to the topic selected in *Topic Navigator* will be popped out in the scatterplot while others are dimmed.

uses a bi-directional barchart representing the accumulative score of positive/negative attitudes in specific intervals to visualize sentiment fluctuations over time, allowing users to identify critical events and how events influence public opinion (DR4). The positive score is encoded by blue bars, while the negative one is encoded by orange bars. Users can identify crucial events by observing abnormal fluctuation and peaks of sentiment bars. Interactive features such as time window sliders help users select specific time ranges with evident sentiment fluctuation and observe the overview(Fig. 3.a), which displays corresponding comments to extract insightful patterns.

Overview Scatterplot. Overview Scatterplot (Fig. 1.C) displays the 2D projection of high-dimensional consumer feedback text embedding in a specific time range selected in the Sentiment Trend Timeline view. The points in the scatterplot are color-encoded based on sentiment labels, and the point size reflects the comment's vote-up score, indicating the number of likes received. This multi-encoded scatterplot allows users to gain a comprehensive understanding of overall comment information at first glance (DR1). Since similar texts are positioned closer in the scatterplot, users can identify potential prevalent opinions by distinguishing specific high-density clusters. Observing the rough proportion of color-encoded points within these clusters provides insights into the sentiment tendency. Conversely, users can also explore non-mainstream individual opinions by investigating outliers. To facilitate user exploration of different combinations of comments, the overview component is equipped with a lasso tool. Through lasso selection, users can flexibly choose comment entities and obtain summaries and mainstream topics of these comments in the Abstract Generator view (Fig. 1.E) and the Topics Navigator (Fig. 1.D), respectively (Fig. 3.b). Meanwhile, for the convenience of obtaining detailed information about a single comment, the view provides a tooltip showing corresponding contents when the mouse hovers over a single point.

Topics Navigator. Topics Navigator (Fig. 1.D) automatically extracts and highlights trending topics from comments, presenting them in a word cloud view to assist users in exploring the consumer discourse landscape (DR3). Utilizing KeyBERT [7] with Maximal Marginal Relevance [1], it identifies the most prominent topics and their sentiment tendency within the dataset. The size of the text in the word cloud indicates the prevalence of a topic among comments in the Overview Scatterplot, and its color represents its sentiment tendency. Upon selecting a topic, corresponding data points in the Overview Scatterplot are popped out, while others are dimmed in grey(Fig. 3.c). This feature enables a focused filter based on the

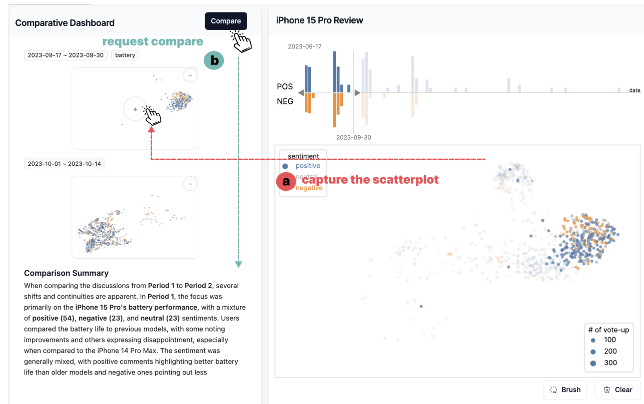

Figure 4: Linking between *Overview Scatterplot* and *Comparative Dashboard*. (a) Users can capture the scatterplot snapshot to *Comparative Dashboard* with the '+' button in the slot. If two snapshots are captured, (b) the 'Compare' button generates *Comparison Summary*.

user's choice of various topics. Furthermore, the navigator updates based on the selected topic, while the word cloud displays prevalent topics calculated only from comments related to the previously selected topic. And users can return to previously selected topics by clicking on blank spaces within the word cloud.

Abstract Generator. Abstract Generator (Fig. 1.E) streamlines analysis of large volumes of consumer comments by distilling them into concise abstracts, thereby providing marketers with an immediate objective estimation of consumer opinions (DR2). This view visualizes sentiment score distributions with easy-to-interpret icons. The abstract is generated by the AI system (GPT-4-Turbo). The important sentences are emphasized through bold text, enabling users to quickly grasp key points in tedious textual content(Fig. 1.E1). Furthermore, Abstract Generator enriches its analysis with a 'Top Comments' list (Fig. 1.E2), which uses a distilled list to prioritize the most impactful feedback. Comments are ranked by their popularity, reflected in the number of likes received, ensuring that the most pertinent and influential consumer opinions are featured prominently. To allow users to locate the position of a specific comment in the Overview, when the mouse hovers over the comment content in the list, the corresponding point will be marked with a red border.

Comparative Dashboard. Comparative Dashboard (Fig. 1.A) serves as a repository to store Overview Scatterplot snapshots of the specific topic in the specific period. It enables direct visual comparison of comments and their dominant sentiment changes over time. Users could use this component to compare pre and post-campaign comments, using coordinated views (Fig. 1.A1) to track clustering and dominant sentiment changes instantaneously. Users can set the comparison periods based on areas of interest or guided by salient peaks identified in the Timeline (Fig. 1.B). Once two snapshots are selected, clicking the compare button can utilize the GPT-4-turbo to generate a summary of the main differences (DR2) and display it at the bottom of the dashboard (Fig. 1.A2) so that users easily understands interesting changes in public attitude and topics.

CommentVis is designed for a user-centric analytical experience, facilitating a natural workflow from macroscopic trends to granular consumer insights. By bridging high-level sentiment trend visualization with the ability to drill down into specific feedback, marketers can craft strategies that are both data-informed and contextually rich.

6 Usage Scenarios

We present interactions and tasks within CommentVis through a scenario featuring Emma, a seasoned marketing professional. Her role involves analyzing customer reactions to products in the smartphone industry. The dataset includes around 2,000 Reddit comments about iPhone 15 Pro, pre-processed by CommentVis backend. Emma utilizes CommentVis to conduct her analysis. iPhone 15 Pro is selected for our usage scenario to demonstrate CommentVis's features.

Identify Key Periods (DR 4). Emma begins by examining the Sentiment Trend Timeline (Fig. 1.B), where bar lengths represent the volume of opinions blue for positive and orange for negative sentiments. Through her experience and understanding of the importance of initial reactions, she notes a significant number of comments following the product launch. To focus on the early feedback, she adjusts the Trend Timeline to the initial post-launch weeks from (2023-09-17 $\sim$ 2023-09-30) by manipulating the graph's slider (Fig. 3.a).

Comprehensive Overview Analysis (DR 1). Emma notices that Overview Scatterplot (Fig. 1.C) is updated by the selected periods. She identifies blue points as positive, orange as negative, and grey as neutral, with point size indicating the number of likes received. By navigating the scatterplot with tooltips like (Fig. 1.C1), she finds that comment points on similar topics seem to be clustered together in the selected period. People mainly discuss Dynamic Island, USB-C, and battery issues. She notes frequent mentions of battery issues.

In-Depth Feature Analysis (DR 3). To understand opinions about topics, Emma chooses the 'battery' (Fig. 3.c) in Topics Navigator (Fig. 1.D). The comment points related to 'battery' popped out in the scatterplot and generated a new set of related topics and abstracts (Fig. 1.E). She finds some negative feedback about the 'battery'. To collate a summary of the feedback, she uses the 'Brush' tool by clicking the 'Brush' button (Fig. 1.C2) to analyze further insights regarding battery in the area. It also prompts the generation of new abstract and negative topics (Fig. 3.b) of a brushed area again.

Investigate User Feedback (DR 2). In Abstract Generator, Emma observes some negative user sentiments by icons. By referring to Abstract (Fig. 1.E1) and Top Comments (Fig. 1.E2), it reveals that users upgrading from models like the iPhone 12 are pleased with the enhancements in speed, quality, and performance. However, some people raised concerns regarding battery longevity, though there is an expectation that the software upgrade will resolve these issues.

Assessing Trend Shifts Over Time (DR 2, DR 4). Emma turns to validate initial observations and assess shifts in Trend Timeline, and she captures the current scatterplot (2023-09-17 $\sim$ 2023-09-30) to the dashboard by clicking the '+' button in the slot (Fig. 4.a). The scatterplot snapshot (Fig. 1.A1) and date label are added in the first slot with the topic. She then modifies the date range to (2023-10-01 $\sim$ 2023-10-14) and adds the snapshot to the second slot. She notices the 'Compare' button is activated; after a few moments of clicking the 'Compare' button (Fig. 4.b), then Comparison Summary (Fig. 1.A2) is generated. She observes a shift in focus from battery issues to an emphasis on battery performance, display quality, and color preferences. The summary implies that battery issues were solved, but to confirm it, she captures the scatterplot topic with 'battery' for each period and compares it again. Comparison Summary confirms that in the second period, the negative sentiment is reduced, and users are more satisfied with battery performance.

In this scenario, Emma efficiently analyzes iPhone 15 Pro feedback with CommentVis in 25 minutes, identifying key sentiments, features, and trends. This method offers quicker, precise insights, outperforming traditional approaches for marketing strategies.

7 User Study

7.1 Objectives and Design

We aimed to assess the practical application and efficacy of CommentVis in real-world marketing analysis scenarios, particularly in extracting meaningful insights from large-scale online user reviews.

Participants. The study involved four experienced marketing professionals from global manufacturing companies: P1 and P2 in product strategy, B1 in brand strategy, and O1 in online business. All participants had 8-10 years of experience. Notably, O1, the first participant, faced difficulties in using CommentVis with only a guide document. Therefore, we excluded O1 from the analysis. This feedback led to the creation of an instructional tutorial video.

Dataset and tasks. The task was to analyze approximately 2,000 Reddit comments, dated from September to December 2023, detailing consumer reactions to Galaxy S23, the latest flagship smartphone manufactured by Samsung Electronics, and expectations for Galaxy S24. Participants were required to analyze reviews, including three primary tasks. These tasks, reflecting design requirements, were designed to be broadly applicable rather than highly specialized and were guided with the assistance of two marketing professionals who helped us define the design requirements. Task 1 (DR1, DR2) is analyzing the comprehensive overview of specific periods. Task 2 (DR3, DR4) is a detailed analysis of the points they're interested in or treated as important periods or features. Task 3 (DR2, DR4) is doing a comparative analysis between topics or trend shifts in periods. Task order and specific methodologies were not prescribed, letting participants conduct as they would in their professional settings.

Procedure. Participants first viewed an 8-minute introduction video including the meaning of CommentVis components, tutorials, and tasks they need to do. The video is provided about a product from a different company to prevent bias from pre-exposure to the dataset and is followed by a 5-minute QA session to clarify any immediate queries. Then, they performed tasks using CommentVis for up to 30 minutes while capturing screens if they accepted, so only P1 and P2 shared the scenario videos additionally. Finally, participants participated in an interview regarding their experience with the tool.

7.2 Results

7.2.1 User Interaction and Observations

Comprehensive Overview Analysis (Task 1). P1 and P2 began their analysis by setting important periods using Trend Timeline (DR4) and then explored Overview Scatterplot. In contrast, B1 started directly with it. All participants were able to identify comment clusters effectively (DR1, DR3). While using Topics Navigator, they identified trending topics (DR3) and assessed the overall user reactions by referring to the generated abstracts (DR2). Notably, they observed a higher positive sentiment, with basic features like battery life and camera capabilities, as well as software aspects such as One UI, Good Lock, and Samsung Dex, receiving praise. These features were often cited as reasons for brand loyalty to Samsung.

In-Depth Analysis (Task 2). Participants conducted in-depth analyses tailored to their professional roles. P1 and P2 focused on period-based (DR4) user reactions to develop timely marketing strategies, while B1 examined product feature reactions to understand product and brand preferences. Utilizing the CommentVis's filtering tools, such as brushes and Topics Navigator, they pinpointed specific responses to features. A notable observation was that all participants easily identified negative opinions about the S24's hardware across various channels, such as dense orange points in the scatterplot and orange topics in Topics Navigator. This finding aligns with our design requirement (DR3). Additionally, the generated topics enabled the participants to understand concerns and related topics about the hardware, facilitating the analysis of further key topics related to it.

Comparative Analysis (Task 3). P1 and P2 effectively utilized Comparative Dashboard to analyze user reactions by time periods or topics. P1, for instance, focused on comparing reactions across various periods. While Trend Timeline and generated abstracts provided individual period reactions (DR4), Comparative Dashboard facilitated a more direct comparison of texts and sentiments (DR2). This tool revealed a notable shift in user focus from S23 reviews in November to the S24 expectations in December, especially following the release of S24's specifications. This trend underscored the simultaneous expression of users' expectations and concerns.

7.2.2 Post-hoc Feedback

Preference. All participants used Overview Scatterplot efficiently. P1 and B1 particularly favored it for its effective guidance on prominent points (DR1). The participants found the clustered points to be an important feature, gaining detailed information through brush and tooltip interactions. Interestingly, Comparative Dashboard was also praised by P1 and P2 for its ease in comparing data (DR2), particularly for topic and period comparisons. However, B1 didn't praise Comparative Dashboard more than P1 and P2 because it was less aligned with her work needs. Trend Timeline was noted for its intuitive usability, closely followed by Topics Navigator (DR3, DR4). Positively, all participants appreciated the efficient linkage and synergy between different views.

Improvements. Suggestions for enhancing CommentVis are mainly related to expanding dataset attributes, such as gender, age, and regions, for more granular filtering. P1 expressed concerns about the clarity of clustering labels; the point echoed in related work [16]. She wished for a clearer understanding of cluster meanings prior to using a brush or Topics Navigator. This aspect is earmarked for future development. B1 suggested the tool could be improved to facilitate broader analysis beyond product features, such as brand image perceptions, not currently displayed in Topics Navigator.

8 Conclusion and Future Work

8.1 Conclusion

Our study established CommentVis as a pivotal tool in the realm of marketing analytics, effectively simplifying the complex task of analyzing consumer feedback. Its innovative visualizations and interactive features, which combine data processing with LLMs, have proven invaluable in extracting meaningful insights from consumer sentiments and discussions on specific attributions of investigated products. The positive feedback from our user study highlights its practical utility in the ever-evolving landscape of digital marketing.

8.2 Future Work

Freeform Navigation. CommentVis is presently limited to preprocessed data, confining exploration within Topics Navigator. Future updates will introduce real-time text embedding processing for user inputs alongside preprocessed data. This work will unlock more dynamic exploration and analysis capabilities within CommentVis, broadening the analytical scope and enhancing user interaction.

Dynamic Clustering To enhance data exploration, we will introduce dynamic clustering tailored to user inputs like topics, inspired by the interactive clustering from Conceptvector [19]. This strategy will enable users to navigate data via clusters aligned with their interests, improving the accuracy and engagement of data analysis.

Generalization. We aim to expand the tool's application across various social media platforms, ensuring its adaptability and effectiveness in a broader range of marketing scenarios.

Scalability. In practical marketing analysis scenarios, there is often a need to deal with a very large size of comments. Additionally, analyzing comments over extended periods is crucial. However, due to the space limitation of visualization, our system may encounter challenges when scaling to a large number of data. The scatterplot and bi-directional bar chart may confront overlap and occlusion, which makes users hard to distinguish clusters and compare sentiment fluctuations. Therefore, improving interaction by incorporating a zoom function to address overlap and occlusion should be considered.

Reliability. While GPT-4-turbo is famous for its good performance in text generation, it is also well-known for generating 'hallucinations': text that is nonsensical or unfaithful to the provided source input [12]. This may result in generated abstracts lacking reliability. Therefore, users should be able to trace back and verify specific parts of the abstract by accessing corresponding comments.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2023R1A2C200520911), and in part by Samsung Electronics. The authors thank Eunsun Cho for suggesting basic requirements. The authors also appreciate Seokweon Jung, Hyeon Jeon and Sihyeon Lee for sharing their valuable knowledges.

References

[1] K. Bennani-Smires, C. Musat, A. Hossmann, M. Baeriswyl, and M. Jaggi. Simple unsupervised keyphrase extraction using sentence embeddings. Data, Analytics & AI, Swisscom AG; Machine Learning and Optimization Laboratory, EPFL, 2018.

[2] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*, 2020.

[3] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, et al. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*, 2023.

[4] W. Cui, S. Liu, L. Tan, C. Shi, Y. Song, Z. Gao, H. Qu, and X. Tong. Textflow: Towards better understanding of evolving topics in text. *IEEE transactions on visualization and computer graphics*, 17(12):2412–2421, 2011.

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[6] S. Faridani, E. Bitton, K. Ryokai, and K. Goldberg. Opinion space: a scalable tool for browsing online comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1175–1184, 2010.

[7] M. Grootendorst. Keybert: Minimal keyword extraction with bert. `https://github.com/MaartenGr/KeyBERT`, 2020.

[8] N. Hu, P. A. Pavlou, and J. J. Zhang. Why do online product reviews have a j-shaped distribution? overcoming biases in online word-of-mouth communication. *Commun. ACM*, 52(10):144–147, 2009.

[9] S. Huang, W. Peng, J. Li, and D. Lee. Sentiment and topic analysis on social media: A multi-task multi-label classification approach. In *Proceedings of the 5th Annual ACM Web Science Conference, WebSci '13*. ACM, New York, 2013.

[10] H. Jeon, A. Cho, J. Jang, S. Lee, J. Hyun, H.-K. Ko, J. Jo, and J. Seo. Zadu: A python library for evaluating the reliability of dimensionality reduction embeddings. In *2023 IEEE Visualization and Visual Analytics (VIS)*, pp. 196–200. IEEE, 2023.

[11] H. Jeon, Y.-H. Kuo, M. Aupetit, K.-L. Ma, and J. Seo. Classes are not clusters: Improving label-based evaluation of dimensionality reduction. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):781–791, 2024. doi: 10.1109/TVCG.2023.3327187

[12] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, and P. Fung. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38, 2023.

[13] S. Lespinats and M. Aupetit. Checkviz: Sanity check and topological clues for linear and non-linear mappings. In *Computer Graphics Forum*, vol. 30, pp. 113–125. Wiley Online Library, 2011.

[14] D. Loureiro, F. Barbieri, L. Neves, L. E. Anke, and J. Camacho-Collados. Timelms: Diachronic language models from twitter. *arXiv preprint arXiv:2202.03829*, 2022.

[15] L. McInnes, J. Healy, and J. Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[16] H. M. J. Mustafa, M. Ayob, D. Albashish, and S. Abu-Taleb. Solving text clustering problem using a memetic differential evolution algorithm. *PLoS ONE*, 15(6), 2020. doi: 10.1371/journal.pone.0232816

[17] A. Neelakantan, T. Xu, R. Puri, A. Radford, J. M. Han, J. Tworek, Q. Yuan, N. Tezak, J. W. Kim, et al. Text and code embeddings by contrastive pre-training. *arXiv preprint arXiv:2201.10005*, 2022.

[18] OpenAI. Gpt-4 turbo. `https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo`, 2023.

[19] D. Park, S. Kim, J. Lee, J. Choo, N. Diakopoulos, and N. Elmqvist. Conceptvector: Text visual analytics via interactive lexicon building using word embedding. *IEEE transactions on visualization and computer graphics*, 24(1):361–370, 2017.

[20] H. Song, Z. Qi, J. Stasko, and D. Yang. Understanding people's needs in viewing diverse social opinions about controversial topics. In *2023 IEEE 16th Pacific Visualization Symposium*, pp. 6–10, 2023.