

Technical Disclosure Commons

Defensive Publications Series

January 2024

Customer Feedback Segmentation, Summarization, and Natural Language Querying Using Machine Learning

Yuanzhi Xu

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation

Xu, Yuanzhi, "Customer Feedback Segmentation, Summarization, and Natural Language Querying Using Machine Learning", Technical Disclosure Commons, (January 29, 2024)
https://www.tdcommons.org/dpubs_series/6641



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

Customer Feedback Segmentation, Summarization, and Natural Language Querying Using Machine Learning

ABSTRACT

It is important for businesses to have an insightful understanding of their customers' needs. It is often difficult to extract statistically meaningful and actionable insights from customer feedback. This disclosure describes the use of machine learning techniques to automatically summarize customer feedback and to generate insights from such summaries using natural language queries. The techniques scale with business size and product volume such that top of mind product issues can be identified from unstructured feedback. By enabling scalable examination, categorization, and natural language querying of customer feedback, the described techniques can help a business discover customer pain points. Raw customer feedback is transformed into actionable intelligence. A user interface can be provided that enables users to provide natural language queries, thus enabling personnel with little or no technical expertise to obtain insights from customer feedback.

KEYWORDS

- Customer feedback
- Text summarization
- K-means clustering
- Prompt engineering
- Large language model (LLM)
- Generative artificial intelligence (Gen AI)
- Natural language query
- Customer review

BACKGROUND

It is important for businesses to have an insightful understanding of their customers' needs. Although customers do leave feedback at business websites, via customer feedback emails, or other mechanisms, it is often difficult to extract statistically meaningful and actionable

insights from such feedback. Furthermore, examining and categorizing feedback can become increasingly time and resource intensive as a business scales and the volume of feedback rises.

While some tools (e.g., [2]) excel in data manipulation and processing, in-memory nature of such tools limits their usability for large datasets. Also, such tools do not inherently support multithreading or distributed computing, which are essential for scalable data analysis. The lack of direct support for concurrency and distribution requires data scientists to seek alternative solutions or additional tools when working with data at scale, adding complexity to the data processing pipeline. Additionally, conversion between data formats is necessary to use deep learning libraries which introduces an additional layer of complexity and can lead to increased memory overhead, thus impacting scalability of data processing for larger and more complex deep learning tasks.

DESCRIPTION

This disclosure describes the use of machine learning techniques to automatically summarize customer feedback and to generate insights from such summaries using natural language queries. The techniques scale with business size and product volume such that top of mind product issues can be identified from unstructured feedback.



Fig. 1: Summarizing customer feedback

Fig. 1 illustrates summarizing customer feedback. The text of customer feedback (102) is transformed to numerical vectors, known as embeddings (106), using a machine learning based

text-to-vector embedder (104). The embeddings capture the semantics of the customer feedback. A clustering module (108) groups semantically similar embeddings into clusters (110). An example technique that can be used by the clustering module is the k-means clustering technique. A large language model (LLM) based summarizer (112) analyzes each cluster and provides a summary of the cluster (114). Summarization enables the identification of common themes and patterns among customer complaints and feedback.

Example of summarization

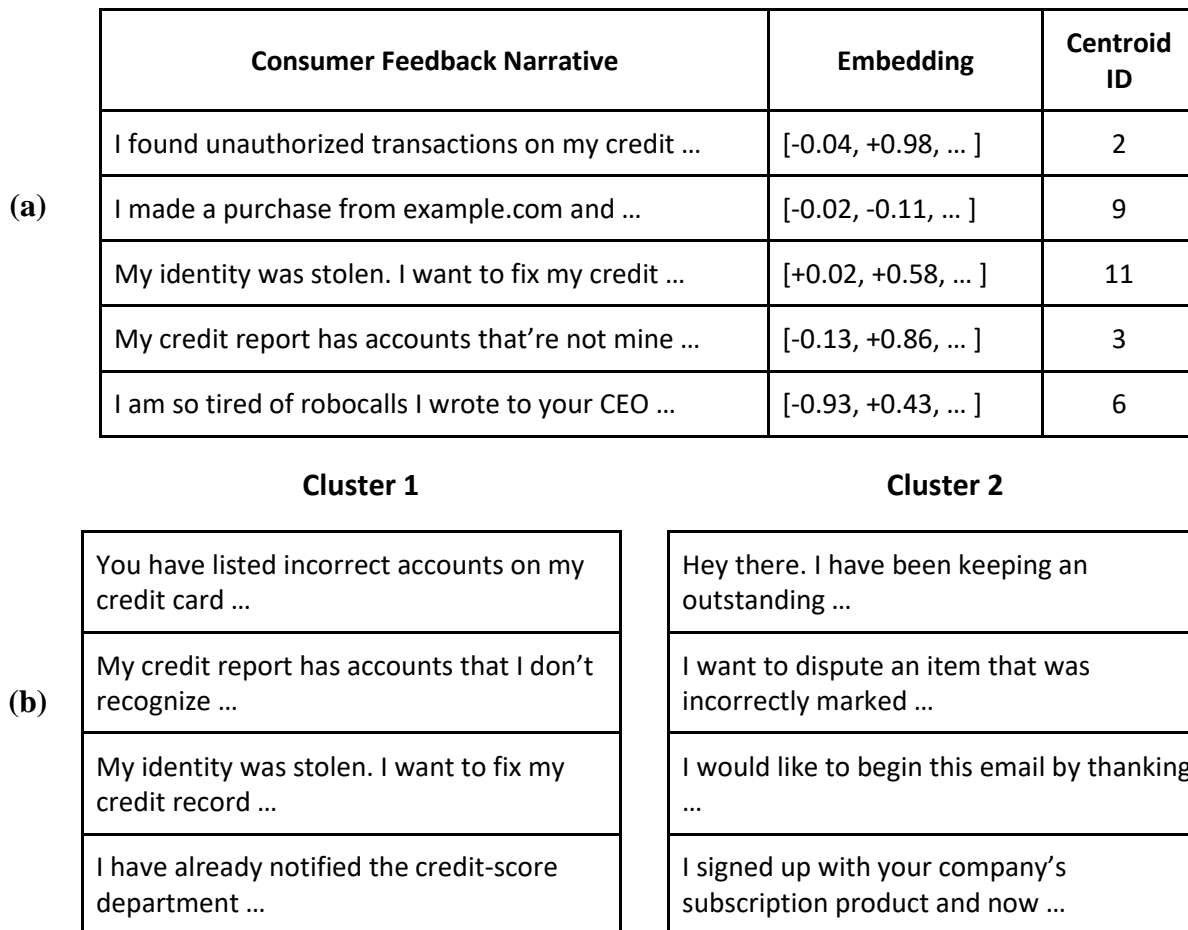


Fig. 2: Summarizing and clustering customer feedback

Fig. 2 illustrates an example of summarizing and clustering customer feedback. Fig. 2(a) illustrates customer feedback narratives (first column) and corresponding embeddings (second column). The third column of Fig. 2(a) identifies the cluster that the feedback is classified into, indicated by the cluster centroid. Fig. 2(b) illustrates examples of clusters, each cluster comprising semantically similar customer feedback.

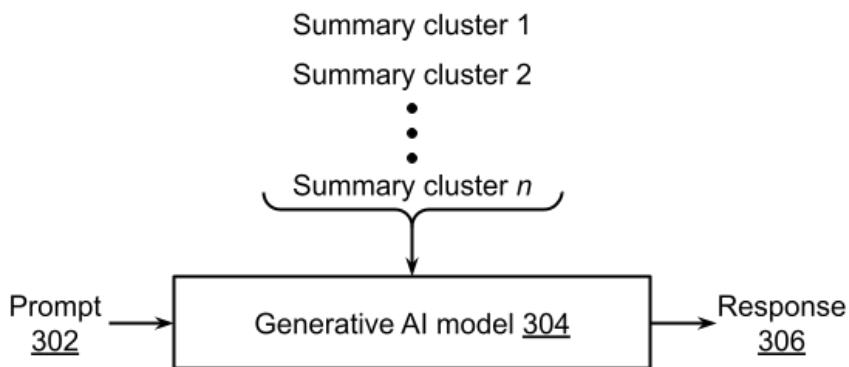


Fig. 3: Natural language querying of customer feedback

Fig. 3 illustrates natural-language querying of customer feedback. The summary clusters are provided as input to a generative AI model (304), along with a user prompt (302). The model generates a response (306). As illustrated in the examples below, actionable insights into the nature of the issues being raised can be obtained using natural language prompts.

Examples of querying using natural language prompts

<p>User-generated prompt: What is the principal issue being raised in Cluster 1? AI-generated response: Cluster 1 mainly deals with identity theft. Prompt: What are the salient differences between Cluster 1 and Cluster 2? Response: Cluster 1 is focused on identity theft while Cluster 2 deals with credit scores. Prompt: What is the most frequent customer complaint of last week? Response: The biggest customer complaint last week was about a web server outage. Roughly 68% of complaints were related to the outage.</p>
--

Since user interaction with the described feedback analysis tool is via natural language prompts, analysis of customer feedback can be performed even by personnel with limited technical skills, enabling such personnel to focus on the business aspects of the enterprise.

The described techniques can be implemented by calling summarization and natural language processing engines from scripts or clients with database querying and programming languages such as SQL and Python and are thereby usable by a wide range of developers. For example, the techniques can be implemented in a distributed (cloud-based) data warehouse with built-in machine learning capabilities and concurrency support. The data warehouse can optimally be accessed by a client that enables interactions with the data warehouse by compiling user code in a programming language such as Python into SQL queries or API calls that are compatible with the data warehouse. Scalability and resource management in machine learning can be handled with a suite of ML modules such as data preparation, data manipulation, model training, model evaluation, model deployment, model serving, etc. provided as part of the data warehouse.

CONCLUSION

This disclosure describes the use of machine learning techniques to automatically summarize customer feedback and to generate insights from such summaries using natural language queries. The techniques scale with business size and product volume such that top of mind product issues can be identified from unstructured feedback. By enabling scalable examination, categorization, and natural language querying of customer feedback, the described techniques can help a business discover customer pain points. Raw customer feedback is transformed into actionable intelligence. A user interface can be provided that enables users to

provide natural language queries, thus enabling personnel with little or no technical expertise to obtain insights from customer feedback.

REFERENCES

1. “CFPB Consumer Complaint Database,” available online at <https://console.cloud.google.com/marketplace/product/cfpb/complaint-database> accessed Dec 10, 2023.
2. “Pandas” available online at <https://pandas.pydata.org/> accessed Dec 10, 2023.
3. “BigQuery DataFrames - Python” available online at <https://cloud.google.com/python/docs/reference/bigframes/latest> accessed Dec 10, 2023.
4. “BigQuery DataFrames” available online at <https://github.com/googleapis/python-bigquery-dataframes> accessed Dec 10, 2023.
5. “Class PaLM2TextEmbeddingGenerator (0.14.1) | Python client library | Google Cloud” available online at <https://cloud.google.com/python/docs/reference/bigframes/latest/bigframes.ml.llm.PaLM2TextEmbeddingGenerator> accessed Dec 10, 2023.
6. “The CREATE MODEL statement for K-means models | BigQuery | Google Cloud” available online at <https://cloud.google.com/bigquery/docs/reference/standard-sql/bigqueryml-syntax-create-kmeans> accessed Dec 10, 2023.
7. “Overview of Generative AI on Vertex AI | Google Cloud” available online at <https://cloud.google.com/vertex-ai/docs/generative-ai/learn/overview> accessed Dec 10, 2023.

8. “Text embeddings | Vertex AI | Google Cloud” available online at <https://cloud.google.com/vertex-ai/docs/generative-ai/model-reference/text-embeddings> accessed Dec 10, 2023.
9. “Text | Vertex AI | Google Cloud” available online at <https://cloud.google.com/vertex-ai/docs/generative-ai/model-reference/text> accessed Dec 10, 2023.
10. “Snowpark ML: Machine Learning Toolkit for Snowflake” available online at <https://docs.snowflake.com/en/developer-guide/snowpark-ml/index> accessed Dec 10, 2023.