LLM Assisted Analysis of Text-Embedding Visualizations

Allen Detmer, Raj Bhatnagar, Jillian Aurisano
Department of Computer Science, University of Cincinnati

Introduction and Motivation

Dimensionality reduction is a widely adopted tool in Natural Language Processing (NLP). Techniques such as Uniform Manifold Approximation and Projection (UMAP) transform high-dimensional embeddings of text data into a lower-dimensional space for visualization. Two-dimensional plots of these embeddings aid in developing insights into model performance. To make sense of these plots, users need to inspect the underlying text represented by the points which can be time-consuming and cognitively intensive. To address this challenge, we developed a novel approach for summarizing and analyzing data behind user selections in text embedding plots. Our interactive approach involves allowing the user to make selections on the text embedding and then utilizing a large-language model (LLM) for: getting a quick overview of the selection, identifying instances of miss-classification, understanding text data within a mixed-class selection, and suggesting additional labels that better fit the underlying text. We implemented our approach in a prototype application, **Text-Embedding Selection Sidekick (TESS),** and present our initial results.

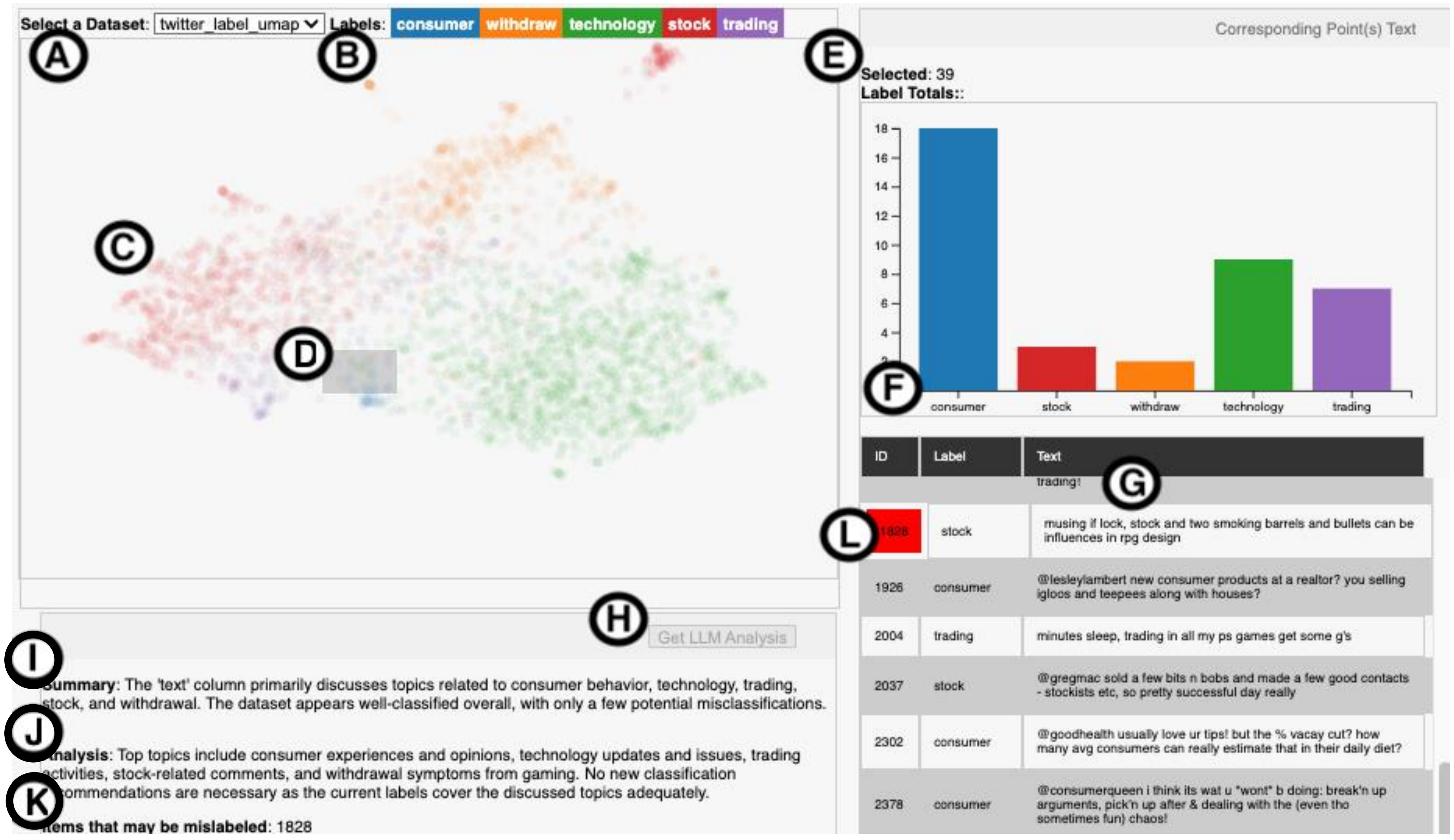
Text-Embedding Selection Sidekick (TESS) Goals, Tasks

1. Reviewing the underlying text data and assessing general model performance
2. Explaining small regions of embedded space having overlapping labels
3. Identifying and explaining mislabeled or misclassified text
4. Identifying classifications that would better partition the data

Data and Pre-processing

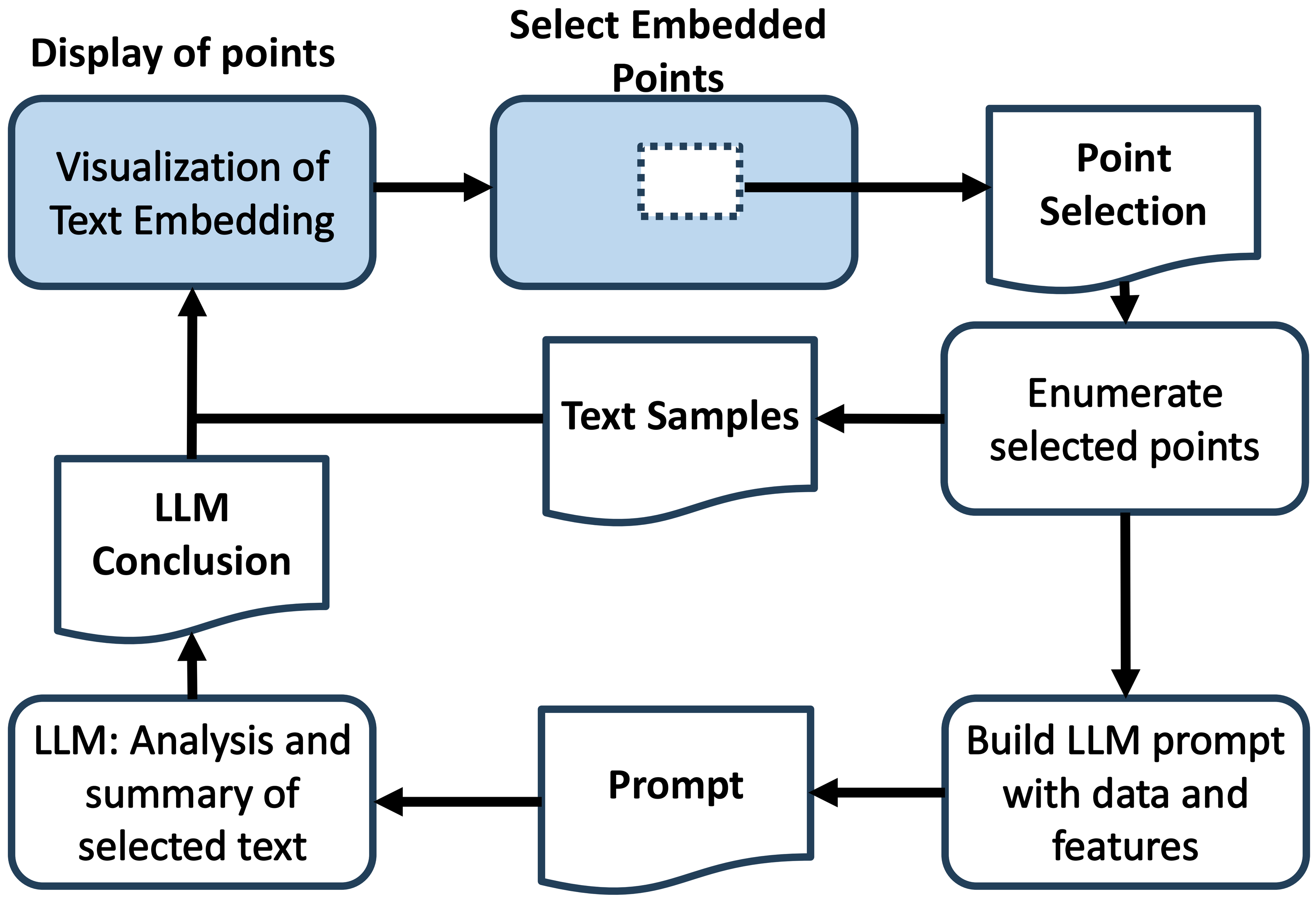
We utilized two multi-class labeled datasets for testing. These datasets were selected because they show different distributions of text similarity and accuracy in labeling, allowing us to explore the use of TESS for model debugging and refinement. The first dataset is emotion-labeled tweets, and the second dataset is tweets potentially related to financial content.

Text-Embedding Selection Sidekick (TESS) Visualization and Interface



A. Dataset selection
B. Text labels in the dataset
C. Two-dimensional text embedding points
D. Brush selection of points
E. Number of selected points
F. Bar chart displaying the count of each label in the brushed selection
G. Table showing identifier, label and text of brushed selection, with sortable columns
H. Button for prompting the LLM driven system to analyze the underlying text and labels of brushed selection data
I. Summary generated by the LLM for the selected data points
J. Analysis generated by the LLM for the selected data points
K. Items that the LLM identifies as possibly misclassified
L. Corresponding items in the table that the LLM identified as possibly misclassified

System Process Flow



The interactive process flow for TESS. TESS displays the embedding and other views of the data utilizing D3. Regions of points are selected with a brush are enumerated and inserted with markdown into a single-shot LLM prompt. The LLM prompt includes instructions to provide a summary, analysis, and possible misclassified labels. The LLM generated output is then displayed in the TESS application interface.

LLM Prompt

 # Dataset Introduction

The following dataset has end user posts in the column 'text' that is classified by the corresponding 'label' representing the classification with a unique identifier in column 'id' for entry: [{table_data}]

 # Task Instructions

Analyze in less than a 100 words the 'text' column for misclassified, identify the top topics discussed in column 'text', and are there new 'text' classification recommendations. Give your final answer in a JSON object format that has two text properties called 'summary' and 'analysis'. A third property called 'misclassified' that is a list of identifiers from the 'id' column of misclassified rows

Conclusion and Future Work

TESS is a prototype tool for exploring interactive visualizations with LLM assisted explanations and analysis. Initial internal tests with TESS suggest such LLM assisted tools could significantly reduce manual and time-consuming text analysis tasks. Future work will involve conducting user studies and evaluations for exploring system effectiveness and future tool enhancements. We also intend to address how our approach might scale to larger datasets and longer text inputs. In internal testing we noted a few instances where the LLM did not consistently follow prompt instructions. We addressed this issue through prompt refinement and verification functions in our system. We will research other techniques eg. one-shot and chain-of-thought (CoT) for improved quality of the LLM response. We will expand techniques for verification of LLM-generated responses in future work.