Deep Learning Final Project Report

Member

Rahul Rangarao - rmr3923
Blake DeLong - bhd343
Yang Wang - yw22729
Kevin Hou - kh37228

Problem Statement

For more than a decade, the barrier to creating and finding reviews about businesses, products, or services has been steadily lowered. Google Reviews, Yelp, Amazon Reviews, Angie's List, TripAdvisor, and similar platforms all act as resources for consumers to make a more informed purchase decision as well as potential sources for businesses to receive consumer feedback. While theoretically valuable to both parties, we believe that most businesses are ill equipped to process the potentially insurmountable amount of natural language data. Specifically, we see a need for businesses to be able to more precisely recognize the negative aspects of their services/products customers are commenting about. Subsequently, businesses would be able to directly respond to negative reviews with this specific knowledge in hand. As a motivation, research has shown that a business's response to negative reviews has a profound positive impact on both customer retention and revenue growth. Specifically, responding to negative reviews increases dissatisfied customer retention rate by 96%, negative reviews that are responded to have a 33% chance to be reversed or removed, and responding to negative reviews increases visibility and ranking in Google's search algorithm. Clearly, if businesses are able to respond to more negative reviews without the heavy investment in training and labor, a great amount of value can be realized.

Proposed Solution

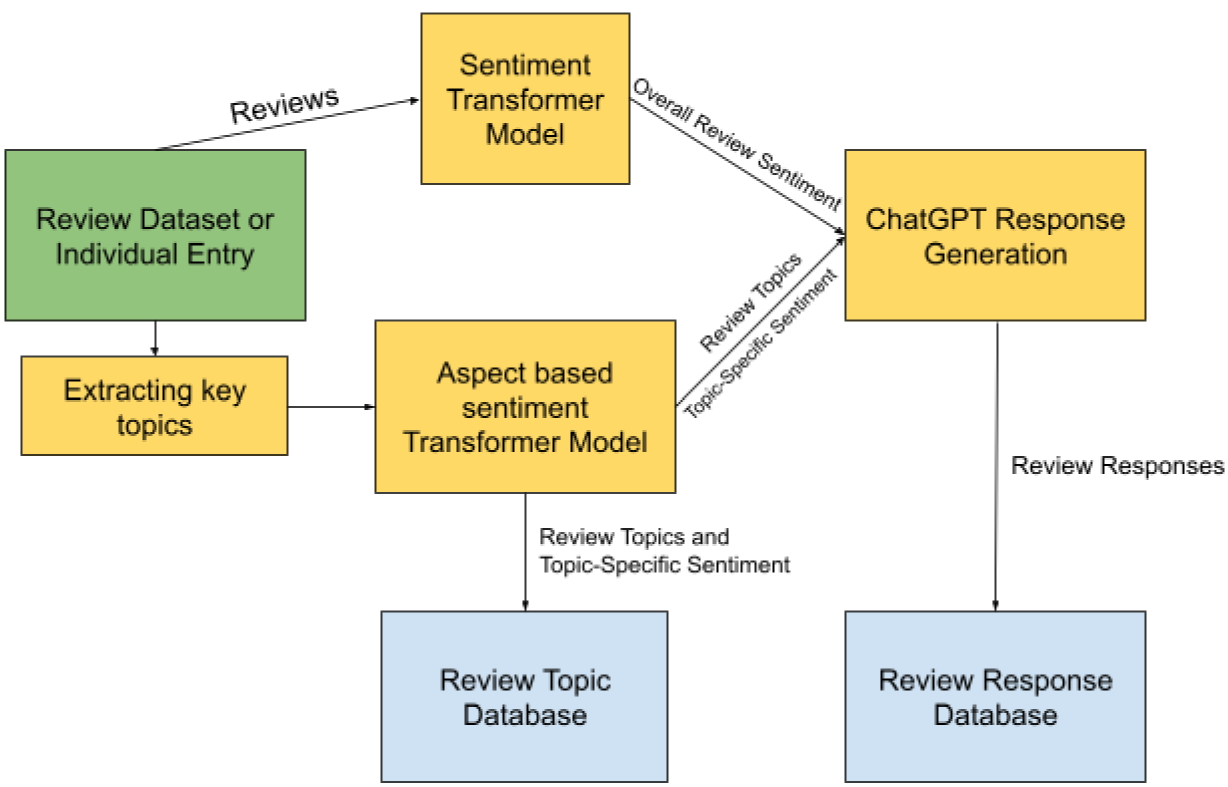
ChatGPT can be a proper tool to help business owners respond to the reviews as it can process and generate responses quickly, saving business owners time and effort, and can provide consistent responses across multiple reviews, ensuring that the tone and messaging are in line with the business's values. To use this tool, the following steps are involved. First, sample data is gathered to train topic modeling and sentiment analysis models. These models are then applied to incoming reviews to determine the overall tone of the review, perform topic extractions, and find out the sentiment of each topic. Next, a proper prompt is created, taking into account the sentiment and topics of the review. This prompt is then sent to the ChatGPT API, which would generate a response based on the information provided. The results would be sent back to the business owner.

Data Source

Initially, we looked to Kaggle to retrieve data sets that contained reviews about a variety of businesses and services. We first started with a review dataset that had review sentiment as a binary positive/negative target label. This dataset was useful in fine tuning the sentiment analysis transformer model, which will be discussed further in the methodology section. In a search for more data, we also retrieved a Kaggel dataset of Yelp Reviews that had additional features related to the reviews such as user feedback on if they found the review useful, funny, or cool. We used these reviews to supplement our existing bank of reviews to analyze.

Methodology

Solution Architecture



Overall Sentiment Analysis:

For this project, we did two sentiment analysis, review sentiments and topic-based sentiment. Both sentiment analysis models are [pre-trained model] for from hugging faces. The pre-trained model, Distilbert-base-uncased-emotion, is fine-tuned to detect the emotion in text. It is faster than any Bert model and with a 93.8 percent of accuracy. For example :

"decide eat here, aware going take 2 hour beginning end. tried multiple times, want like it! location NJ never bad experience. food good, take long time come out. waitstaff young, usually pleasant. many experience spent way long waiting. usually opt another diner restaurant weekends, order done quicker."
[{'label': 'POSITIVE', 'score': 0.9840111136436462}]

"second time tried turning point location. first time long wait food ordering, time even longer wait 40 minutes. omelette skillet hardly egg it, felt like eating chopped onion chopped tomatoes. wife BLT hard time finding tomato avocado supposed it. Overall,The experience stressful, mainly long wait."
[{'label': 'NEGATIVE', 'score': 0.9976387023925781}]

The overall sentiment scores give us an overall emotion of the review which can help us improve the performance of the review response generation. The label will be passed to the gpt model as a parameter to make the model generate a more accurate response. This is a part of prompt engineering.

Individual Topics and their sentiments:

With this aspect we wanted to perform two things: Identify the "key topics" within each review and give an overview of the business of what customers are talking about and their individual topic sentiments. For example in the review:

"If you decide to eat here, just be aware it is going to take about 2 hours from beginning to end. We have tried it multiple times, because I want to like it! I have been to it's other locations in NJ and never had a bad experience.

The food is good, but it takes a very long time to come out. The waitstaff is very young, but usually pleasant. We have just had too many experiences where we spent way too long waiting. We usually opt for another diner or restaurant on the weekends, in order to be done quicker."

We can see that the review talks about "Wait time", "Food", "Service" tags. The other tags mentioned in reviews were "Ambience" or "Amenities"

- BERTopic: BERTopic is a topic modeling technique that leverages transformers and c-TF-IDF to create dense clusters allowing for easily interpretable topics whilst keeping important words in the topic descriptions. However, the issue with this is that this is usually done with documents that are usually non-homogenous and our review dataset mainly homogenous.

The resulting topics (Top 5) were:

- ★ food_place_order
- ★ food_place_good_great
- ★ yelp_review_food_place
- ★ reno_food_sabrina_great
- ★ ihop_mex_tex_tex mex

Topic 1 talks about the overall place along with the order experience, Topic 2 talks about the food and the excellent experience that the customer had. Topic 3 relates to yelp reviews or reviews and their impact on businesses. Topic 4 is a specific reno based food chain and Topic 5 is about iHop compared to their Tex-Mex counterparts

We can see that this did not perform a good job in summarizing and segment the reviews. Our expectation was to get more nuanced attribute based topics such as “Food quality” or “Service time” or “Overall location or Ambience”. We therefore tried an alternate technique of summarization

- Review summary and attribute mapping

The first aspect of this method is to summarize the reviews. Within summarization there are two techniques, such as extractive summarization that works by extracting several parts, such as phrases and sentences, from a piece of text and stack them together to create a summary, and abstractive summarization which utilizes advanced NLP techniques to generate an entirely new summary. The abstract summarization uses a technique known as the “TextRank” algorithm that uses graph based ranking model for text processing. The “TextRank” algorithm is similar to the page rank algorithm employed by Google. Gensim has an implementation of this algorithm in their summarization module

Post obtaining the summaries from the review, we then map it using one of the attribute word using spacy similarity. The attributes and their keywords are:

"food": ["food", "cuisine", "dish", "flavor","fresh"],
"service": ["service", "waitstaff", "server", "staff","manager"],
"ambiance": ["ambiance", "atmosphere", "decor", "music","place","seat"],
"value": ["value", "price", "cost", "affordability"],
"location": ["nearby", "address", "parking", "accessibility","locality","neighbourhood"]

This was performed on a set of ~1000 reviews

Post obtaining the key topics, we then had to perform individual sentiment analysis of each topic. Going back to the review:

"If you decide to eat here, just be aware it is going to take about 2 hours from beginning to end. We have tried it multiple times, because I want to like it! I have been to it's other locations in NJ and never had a bad experience.

The food is good, but it takes a very long time to come out. The waitstaff is very young, but usually pleasant. We have just had too many experiences where we spent way too long waiting. We usually opt for another diner or restaurant on the weekends, in order to be done quicker."

The above review talks about **"wait time"**, **"service"** and **"food".** This was then fed into Aspect base sentiment analysis

Aspect-Based Sentiment Analysis (ABSA), also known as fine-grained opinion mining, is the task of determining the sentiment of a text with respect to a specific aspect. Aspect-based sentiment analysis is a relatively new field of research that has emerged in response to the limitations of traditional sentiment analysis methods.
The technique was introduced by researchers from University of Beijing and University of Exester. In their paper, they highlight the need for a more nuanced approach to sentiment analysis and identifying key topics in a review and their corresponding sentiments.

This uses a supervised technique and reviews need to be manually updated as negative positive or neutral.

ID	Text	Overall
1	If you decide to eathere, just be aware it is going to take about **2 hours** from beginning toend. We have tried it multipletimes, because I want to likeit! I have been toit's other locations in NJ and never had a badexperience. The **food** isgood, but it takes a very long **time** to comeout. The **waitstaff** is veryyoung, but usuallypleasant. We have just had too many experiences where we spent way too long **waiting.** We usually opt for another diner or restaurant on theweekends, in order to be donequicker.	Negative

Since this would take time, we used a pre-trained dataset, to get the custom embeddings. There is also a pytorch implementation to train the dataset in GitHub. This was used for our implementation purposes. The supervised technique has an accuracy of **84%** and the F-1 score is around **76%**
The output of the above review with the topics **"wait time"**, **"service"** and **"food quality"** is:

{
'Wait time': {'Negative': 0.95},
'Food quality': {'Positive': 0.84},
'Service': {'Negative': 0.72}
}

Response Generation Solution

Now that overall review sentiment has been established, the review has been summarized to its core topics, and Aspect-Based Sentiment Analysis has been applied, we now look to use these outputs to create a prompt for ChatGPT. To start, we decided to use the ChatGPT API in order to create a more automated solution that would be able to handle large datasets. Essentially, the API is very similar to the web-app in that a user prompts the model and the model returns a response. The only difference is that each API call requires all previous user/model interactions if a user wishes to carry on a conversation. For our purposes, we do not need to query the model more than once per review, so logging conversations was not necessary.

To begin building our prompt, we first have to consider if a response is even needed. In other words, we created a small function that takes the overall sentiment of the review and returns a boolean value if we determine that the review is negative enough to respond to. The output of this function then determines if we move forward to create the prompt for the API call.

Using f-strings, we are then able to create an outline of the prompt we would like to use and then inject the topics and sentiments we gathered from the topic modeling and aspect-based sentiment transformer model. The "skeleton prompt" looks like this:

You are an apologetic business owner and your purpose is to respond to negative reviews. You will be given topics that the customer mentioned and the associated sentiment score in the form of [Topics: {sentiment:score}].

Your objective is to respond to the reviews effectively. You can choose to ignore certain topics or overall sentiment, if you deem them to be unnecessary. Do not offer a discount. Mention that the customer can reach out to management at {email}. Do not exceed 5 sentences.

An example of topics and sentiment that would be injected into the prompt are as follows:

[Food:{Positive:0.99}, Waitstaff:{Negative:0.7}, Manager:{Negative:0.8}, {Ambience: Neutral:0.5},Value For Money:{Positive:0.8}, Freshness: {Positive:0.1}]

Additionally, the user can choose to enter a contact email address for the reviewer to reach out to.

Using this example, ChatGPT generates this response:

Thank you for sharing your feedback with us. We're sorry to hear that you were not satisfied with the service you received from our waitstaff and management team. We strive to provide excellent *service to all of our customers, and we're disappointed to hear that we fell short in this instance. We will take steps to address these issues with our team to ensure that we improve our service moving forward. We're glad to hear that you enjoyed our food and found it to be a good value for money. We appreciate your business and hope that you'll consider giving us another chance. If you have any further concerns, please feel free to reach out to us at blakedelong@utexas.edu.*

The response is then provided to the user for final editing and approval.

Reflections

Our project aimed to develop a review management suite that would help businesses respond to negative reviews in a more efficient and effective manner. Overall, we are pleased with the response that we received from our review management suite. Our tool was able to quickly and accurately identify negative reviews, extract key topics, and generate appropriate responses. We believe that our tool can significantly reduce the time and cost associated with responding to negative reviews while also creating business value.

One of the unexpected use cases of our project was the use of topic modeling to summarize the key themes of negative reviews. As outlined, our tool is able to extract the most important topics in reviews in order to use them for prompt generation. However, these topics can also be collected in bulk and fed into a single ChatGPT prompt, asking the model to summarize what reviewers have to say about this business. Essentially, this allows ChatGPT to summarize hundreds of reviews that would normally exceed the token length. This summarization could prove useful to upper level management that needs to see a broad summary of what is being discussed.

However, we recognize that our project is still in its proof of concept stage and would need to be refined to be seamlessly integrated into a social media/review manager's workflow. We believe that further development and testing would be needed to ensure that our tool can be effectively used by businesses of varying sizes and industries.

Despite these limitations, we are confident that our project has the potential to be a valuable tool for businesses seeking to improve their response to negative reviews. We hope that our project will inspire further research and development in this area and ultimately contribute to improving customer satisfaction and retention.

In conclusion, our project successfully demonstrated the potential of using deep learning and natural language processing techniques to develop a review management suite that could help businesses respond to negative reviews more efficiently and effectively. We are proud of the progress we made and believe that our work has the potential to make a real-world impact.