

HOTEL REVIEW RECOMMENDER

Yi-Hung, Lu
American University

ABSTRACT

This research applies Sentence-BERT, prompt engineering, and large language models (LLMs) to generate summaries. The study compares the effects of dataset size on text generation using two different LLM models.

Keywords: NLP, Text Generation

1. INTRODUCTION

Travelers today are inundated with online reviews and ratings, which can aid decision-making but also lead to decision fatigue. Limited time prevents users from reviewing all options, making it hard to find the best match. This growing challenge underscores the need for recommendation systems to streamline data and deliver personalized choices.

2. RELATED WORK

Reimers and Gurevych (2019) introduced Sentence-BERT (SBERT), a method for computing sentence embeddings using Siamese BERT networks. Their approach calculates the similarity between two sentences using cosine similarity. According to their findings, SBERT often outperforms other methods in terms of accuracy for various tasks.

Gao et al. (2024) propose a framework called Retrieval-Augmented Generation (RAG) combines generative and retrieval techniques to improve large language models (LLMs). By utilizing outside information sources during the generation process, this method allows the model to generate responses that are more precise, pertinent to the context, and current.

3. METHODOLOGY

3.1 SENTENCE BERT

The Sentence-BERT (SBERT) model, specifically the pre-trained "all-mpnet-base-v2" variant, is selected as the foundational embedding technique because its effectiveness in producing semantically meaningful sentence representations. Each dataset undergoes fine-tuning of the SBERT model to optimize document embeddings tailored to the specific domain. Finally, the fine-tuned embeddings are utilized to retrieve relevant reviews, enhancing the precision of downstream tasks. Cosine similarity is computed between the question embedding and each review embedding to retrieve relevant reviews. This approach ensures that the most semantically similar reviews to the query are identified, enhancing the relevance and precision of the retrieval process.

3.2 PROMPT ENGINEERING

First, define the query clearly and provide the relevant reference text, such as reviews, to assist the model in answering the query. Instruct the model to generate a response based on the query and the information contained within the reference text. This ensures that the model's answer is grounded in the provided context, improving the accuracy and relevance of the response.

3.3 LLM

The research will utilize two large language models (LLMs): Llama 3.1-70B (address: 'meta-llama/Llama-3.2-1B-Instruct') and ChatGPT with the API (GPT-4 mini). Each model generates a summarized response based on the combined reviews, providing concise and insightful answers to the original query.

4. EXPERIMENTATION AND RESULT

4.1. DATASETS

GitHub-provided dataset, including star ratings, hotel names, review titles, and review texts. The Vienna dataset includes 20301 observations, and the Aspen dataset includes 2006 observations.

	stars	hotel	title	content
0	5	[The Ring Hotel - Relais & Chateaux]	[Excellent location with beautiful spacious ro...]	[It was our first stay in this nice hotel. We ...]
1	5	[Boutiquehotel Stadthalle]	[Excellent service]	[This hotel is amazing. Maybe the best in Vien...]
2	5	[Austria Trend Parkhotel Schoenbrunn Vienna]	[Dancing in Vienna]	[A Once in a Lifetime experience to Waltz in V...]
3	5	[Austria Trend Parkhotel Schoenbrunn Vienna]	[Christmas Markets]	[Great location with access to transportation ...]
4	5	[The Ring Hotel - Relais & Chateaux]	[Great location and nice room!]	[There were four people staying in this room,...]
...
20296	5	[The Guesthouse Vienna]	[Outstanding boutique hotel in centre of Vienna]	[This hotel is perfectly positioned 3 minutes ...]
20297	5	[The Guesthouse Vienna]	[Wonderful stay in the heart of the city]	[What a place! From the warmest of welcomes, g...]
20298	5	[The Guesthouse Vienna]	[Great Recommendation For Breakfast]	[When we were displeased with breakfast option...]
20299	5	[Boutique Hotel Donauwalzer]	[Excellent value]	[We stayed at the Donauwalzter for 2 nights an...]
20300	3	[Boutique Hotel Donauwalzer]	[Great quality price ratio]	[The staff is very friendly. The breakfast is...]
[20301 rows x 4 columns]				

Fig.1. Vienna dataset

	stars	hotel	title	content
0	5	[Limelight Hotel]	[Awesome visit]	[Went on a girls trip this past weekend. We ha...]
1	5	[Limelight Hotel]	[Super hotel and Super Staff]	[We were very lucky to win 4 nights accomdati...]
2	5	[Chateau Roaring Fork]	[Wait until the last minute]	[Wait until the last minute and get a lodging/...]
3	5	[Limelight Hotel]	[Great Hotel. Nice place to stay]	[Great hotel. Beautiful. Great well decorated ...]
4	3	[Aspen Mountain Lodge]	[Quaint and cozy lodge]	[Great value, decent location. I'd highly reco...]
...
2001	5	[Chateau Blanc]	[Great stay]	[Spent a couple of nights in Aspen on a girls ...]
2002	4	[Chateau Blanc]	[Excellent cost/benefit]	[We stayed in a two bedrooms/bathrooms apartme...]
2003	5	[Chateau Blanc]	[Great Winter Vaca]	[A wonderful place to stay for our family vaca...]
2004	5	[Chateau Blanc]	[Chateau Blanc for a week]	[The lodge is few blocks away from the main do...]
2005	5	[Shadow Mountain Lodge]	[Excellent Value]	[Great location, nice rooms much larger than a...]
[2006 rows x 4 columns]				

Fig.2. 5 Aspen dataset

4.2. EXPERIMENT SET UP

4.2.1 DOCUMENT EMBEDDING

The data preprocessing step involves applying BERT tokenization to prepare the text for analysis, followed by generating document embeddings using the Sentence-BERT (SBERT) model. To retrieve relevant reviews, cosine similarity is applied between the question embedding and each review embedding, ensuring the identification of the most semantically similar reviews.

This is top 10 relevance review with query (What are the best views for travel?)

Rank 1: Review 1382 (Score: 0.5003)
Hotel: ['Hotel Jerome']
Review Content: ['An old hotel full of history and character - if you're eating next door you should wander through this ho
Rank 2: Review 1980 (Score: 0.4990)
Hotel: ['St. Moritz Lodge & Condominiums']
Review Content: ['We absolutely loved our fall stay at St Moritz. The location is so convenient to great food and drink. Th
Rank 3: Review 781 (Score: 0.4961)
Hotel: ['Aspen Mountain Lodge']
Review Content: ['Even with man made snow the photos are stunning, and you can relax in a pure beauty location unlike any o
Rank 4: Review 529 (Score: 0.4958)
Hotel: ['Aspen Alps Condominium Resort']
Review Content: ['We had an incredible time at the Aspen Alps-the location is great for summer or winter and very close to
Rank 5: Review 1093 (Score: 0.4943)
Hotel: ['Tyrolean Lodge']
Review Content: ['Quaint hotel just on the edge of town so ideal location for a relaxing walk into town.', 'Room was a good
Rank 6: Review 1962 (Score: 0.4933)
Hotel: ['Aspen Meadows Resort']
Review Content: ['We found the hotel through a work trip, and were surprised to discover an immaculate 40 acre art mecca i
Rank 7: Review 184 (Score: 0.4890)
Hotel: ['Mountain Chalet Aspen']
Review Content: ['We stayed at Mountain Chalet in early October and chose it primarily because the price was attractive. W
Rank 8: Review 1989 (Score: 0.4985)
Hotel: ['Aspen Meadows Resort']
Review Content: ['I visited the Meadows as part of professional training. It was a wonderful resort. The staff is phenom
Rank 9: Review 284 (Score: 0.4814)
Hotel: ['The St. Regis Aspen Resort']
Review Content: ['This hotel is truly exceptional. The room is nicely laid out and includes lush blankets and an exceptiona
Rank 10: Review 1239 (Score: 0.4793)
Hotel: ['Hotel Jerome']
Review Content: ['Was in Aspen for day and so did a walk about and found the Jerome. Amazingly beautiful hotel with great b

Fig.3. Retrieve relevant review

4.2.2 Prompt

The process begins by formulating a query for the recommender system. For example, a typical query could be, "Which hotel has the best room view?" To retrieve relevant reviews, a prompt is developed based on the query. Using document embeddings, the system identifies three relevant reviews. These reviews are then combined to create a "super prompt," which includes the hotel name and the content of each review. This approach ensures that the most pertinent evaluations are effectively summarized to answer the query.

Instruction	Help me to summarize the prompt and respond to our travel-related query. Give me the top three hotels that are suitable for the query. query: 'What are the best views for travel?'(Example: Based on your question, 'What are the best views for travel?', here's the summary: 'hotel_name' consistently receives praise for its remarkable views, particularly of the river. Guests highlight that it has the best scenic vantage points, making it ideal for those seeking breathtaking views.)
Relevant review	Rank 1: Review 1382 (Score: 0.5003) Hotel: ['Hotel Jerome'] Review Content: ['An old hotel full of history and character - if you're eating next door you should wander through this hotel before you head out / full of charm and friendly staff. There's a great huge fireplace at the bar too'] Rank 2: Review 1980 (Score: 0.4990) Hotel: ['St. Moritz Lodge & Condominiums'] Review Content: ['We absolutely loved our fall stay at St Moritz. The location is so convenient to great food and drink. The warm pool (90°F) was a treat. This lodge is clean and cheerful and just chock full of character.'] . . . Rank 10: Review 1239 (Score: 0.4793) Hotel: ['Hotel Jerome'] Review Content: ['Was in Aspen for day and so did a walk about and found the Jerome. Amazingly

	beautiful hotel with great bars and restaurants. I ate in the front bar and then took a tour through hotel. I will have to add this hotel to my bucket list to come back and stay. People were all super friendly and helpful. The foyer is beautiful!"]
--	--

Table.1. super prompt

4.2.3 LLM SUMMARIZING

The "super prompt" is input into two different large language models (LLMs) for processing.

	Result for small dataset (Aspen)
Open AI	<p>Based on your query, 'What are the best views for travel?', here's the summary of the top three hotels that are known for their remarkable views:</p> <ol style="list-style-type: none"> 1. **Hotel Jerome** - This historic hotel is celebrated for its charm and character. Guests appreciate the scenic ambiance, particularly the beautiful fireplace area, which creates a cozy and visually stunning setting. Its picturesque design and surroundings make it an ideal choice for those seeking breathtaking views. 2. **Aspen Meadows Resort** - Renowned for its 40-acre grounds filled with art and natural beauty, this resort boasts stunning views and a unique environment. The stunning architecture, serene hiking trails, and outdoor seating areas allow guests to fully enjoy the breathtaking scenery surrounding the property. Reviewers consistently highlight the immaculate grounds and the scenic views available throughout the resort. 3. **Aspen Mountain Lodge** - Visitors to Aspen Mountain Lodge are captivated by the picturesque views of the surrounding mountains, especially when covered in snow. The lodge offers a beautiful setting for relaxation, paired with essentials for a memorable stay, making it a delightful choice for those looking to soak in mountain vistas. These hotels not only provide comfortable accommodations but also ensure that guests experience some of the most beautiful views available in the region.
Limma	<p>Based on the query 'What are the best views for travel?', here's a summary: The top three hotels that consistently receive praise for their remarkable views are:</p> <ol style="list-style-type: none"> 1. **Hotel Jerome**: Consistently receives high praise for its breathtaking views, particularly of the river, making it an ideal choice for those seeking breathtaking views. 2. **St. Moritz Lodge & Condominiums**: Located in the heart of St. Moritz, this hotel offers stunning views of the surrounding mountains and is a great base for exploring the town and its surroundings. 3. **Aspen Mountain Lodge**: Even with man-made snow, the photos taken here are stunning, and the location is pure beauty, making it a great choice for those seeking a unique and breathtaking view. These hotels are consistently praised for their exceptional views, making them ideal choices for travelers seeking breathtaking views while on the road.

Table.2. small dataset summarizing (Aspen)

	Result for long dataset (Vienna)
--	----------------------------------

OpenAI	<p>**Summary:** Based on your question, 'What are the best views for travel?', here are three hotels that consistently receive praise for their remarkable views:</p> <ol style="list-style-type: none"> 1. **Radisson Blu Palais Hotel, Vienna**: This hotel is lauded for its excellent location, particularly the coffee shop that offers prime vantage points for watching the city below. Guests enjoy a leisurely experience with stunning views as they sip their beverages. 2. **Hotel Am Parkring**: Guests highly recommend this hotel's amazing views, especially from rooms on the side of the restaurant. The location is convenient, with a tourist bus stop right outside, making it an ideal base for sightseeing, all while enjoying picturesque scenery. 3. **Grand Ferdinand**: Notable for its rooftop pool, this hotel offers a panoramic view of the city that guests describe as alluring. It's a perfect spot for relaxation with a refreshing cocktail while taking in the scenery. These accommodations are perfect for travelers seeking breathtaking views during their stay.
Limma	<p>Based on your travel-related query, I have summarized the top three hotels that are suitable for breathtaking views:</p> <ol style="list-style-type: none"> 1. **Radisson Blu Palais Hotel, Vienna** - This hotel consistently receives praise for its remarkable views, particularly of the river. Guests highlight that it has the best scenic vantage points, making it ideal for those seeking breathtaking views. 2. **Hotel Lamee** - Located in the heart of Vienna, this hotel offers top-notch views of the city, particularly from its rooftop terrace. Staff are approachable and friendly, making it a great choice for travelers. 3. **Hotel Gallitzinberg** - This hotel is situated in a quiet area within walking distance to the Vienna Woods and has bus stops at the front door, making it an ideal location for sightseeing. Its rooftop pool and panoramic views make it a great choice for those seeking a relaxing getaway. These three hotels consistently receive high praise for their breathtaking views, making them ideal for travelers seeking stunning vistas while on the go.

Table.3. large dataset summarizing (Vienna)

4.3. RESULT

This research compares the effects of dataset size on text generation using two LLM models. Based on the result, OpenAI's model is expected to generate more adjectives to describe reasons and utilize longer sentences to answer questions. In contrast, Limma, a smaller LLM, tends to use shorter sentences and similar adjectives in its responses. Both models provide understandable and accurate information to address queries.

5. DISCUSSION

An evaluation form should be created for customers to determine their preferences

regarding which model is better. Future research can explore the application of other large language models for summarization, such as Limma - 70B. Additionally, future work can update the dataset to develop a travel recommender system, as customers seek to experience the best and latest travel suggestions rather than outdated ones.

6. CONCLUSION

It is anticipated that the larger dataset used with ChatGPT will yield better results. Ultimately, both OpenAI and Limma generate sensible answers to the queries. Additionally, utilizing the OpenAI API incurs costs, although using the 4o mini version is relatively inexpensive. In contrast, Limma does not require any financial investment. Therefore, when considering cost-effectiveness, Limma may be a more advantageous model to use.

7. REFERENCES

Datasets Address:
<https://github.com/youngjeong46/travel-recommender/tree/main/data/raw>.

Reimers, N., & Gurevych, I. (2019). Sentencebert: Sentence embeddings using Siamese Bert-Networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*.
<https://doi.org/10.18653/v1/d19-1410>

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2024, March 27). Retrieval-augmented generation for large language models: A survey. arXiv.org. <https://arxiv.org/abs/2312.10997>