SENTIMENT ANALYSIS ON HOTEL REVIEW

Yi-Hung Lu, Ran Zhang
American University

ABSTRACT

This essay explores the concept of sentiment analysis in the context of hotel reviews. Sentiment analysis is a valuable tool for understanding customer opinions and satisfaction levels, particularly in the hospitality industry where customer experience is paramount. Using a dataset of hotel reviews, we discuss the process of sentiment analysis, including data preprocessing, feature extraction, and machine learning model implementation. We also examine the challenges and limitations of sentiment analysis in hotel reviews, such as the nuances of language and the impact of fake or biased reviews. Through this exploration, we aim to highlight the importance of sentiment analysis in improving customer service and business strategies in the hotel industry.

1. INTRODUCTION

Hotel reviews are now an essential element of contemporary travel, providing a wealth of information for both travelers and hospitality providers. In today's digital era, travelers can utilize numerous online platforms to share their experiences and opinions regarding hotels, resorts, and other accommodations. These reviews assist future travelers in making informed decisions and provide valuable feedback to hoteliers, enabling them to enhance their services and improve the overall guest experience.

The proliferation of online review platforms has democratized the travel industry, empowering travelers from diverse backgrounds and budgets to express their views. This democratization has ushered in an era of transparency and accountability in the hospitality sector, compelling hotels to prioritize customer satisfaction and pursue excellence in all aspects of their operations.

Within this context, the analysis of hotel reviews has emerged as a valuable tool for travelers and hoteliers alike. By scrutinizing and interpreting large volumes of reviews, researchers and industry experts can uncover trends, patterns, and insights that inform strategic decision-making and drive ongoing improvement. Whether identifying emerging guest preferences or pinpointing areas for operational enhancement, the analysis of hotel reviews presents numerous opportunities for industry stakeholders to gain a competitive advantage and deliver exceptional guest experiences.

2. RELATED WORK

Natural language processing research on travel reviews covers a wide range of subjects and approaches to understanding, evaluating, and producing travel-related information. Sentiment analysis is a basic approach to comprehending information that comes from travel reviews. Researchers already explored different techniques regarding machine learning models, deep learning models, and transformer-based models to classify the sentiment expressed in travel reviews as positive, negative, or neutral.

Authors try to use the BERT model with travel reviews for several reasons including context understanding model performance and transfer learning. Chu et al. (2022) demonstrated, “This work proposes an aspect-based sentiment analysis model by extracting aspect-category and corresponding sentiment polarity from tourists' reviews, based on the Bidirectional Encoder Representation from Transformers (BERT) model.” Compared to deep learning, in this study, Martín et al. (2023) found that the LSTM model applying travel reviews was better than other models in deep learning models. The LSTM model provides accurate and robust estimators. However, the LSTM model relies on sequential processing, making it potentially challenging to capture information as the BERT model. Moreover, the LSTM model will spend more time than the BERT model because it needs to require more data to train from scratch compared to the BERT model's pre-training on large text corpora. Overall, this article expected that the BERT model could provide more accurate sentiment analysis with travel reviews.

3. METHODOLOGY

The approach involves preprocessing the reviews to calculate TF-IDF scores, identifying key terms indicative of sentiments and important features. Subsequently, the preprocessed reviews are tokenized and fed into BERT to generate contextual embeddings, capturing nuanced sentiment expressions. A sentiment analysis model is then trained using these embeddings, and its performance is evaluated in predicting sentiment labels. The methodology also includes the interpretation of TF-IDF features and BERT embeddings to understand sentiment indicators and classify new reviews, providing valuable insights into customer opinions. Optionally, multilevel models can be used to predict hotel ratings based on sentiment labels, further enhancing the analysis. Overall, this methodology aims to improve sentiment analysis on hotel reviews, offering deeper insights for decision-making and enhancing customer satisfaction.

3.1. TF-IDF

By applying TF-IDF to hotel reviews, it extracts keywords and key phrases that are characteristic of positive or negative sentiments, specific aspects of the hotel experience or other important features mentioned in the reviews. These extracted features can then be used for sentiment analysis, topic modeling, or other tasks to gain insights from the reviews. Based on TD-IDF feature, it is calculated as the logarithm of the total number of documents divided by the number of documents that contain the term. “The smaller this probability, the less probable it is that the text got k occurrences randomly, and thus, the more confident we are that the word t is important for the given document.” (Havrlant and Kreinovic, 2017) It will get TF-IDF score for a term in a document is the product of its TF and IDF scores. This score helps identify terms that are both frequent in the document and unique to it, making them more informative for understanding the review's content.

3.2. LSTM

To build model, the LSTM as part of RNN have to be used. The LSTM's ability to remember long-term dependencies can help it capture subtle nuances in language, allowing it to accurately classify the sentiment of each review. The text generation and summarization rely on those subtle nuances to generate the new sentences. By training the LSTM on a corpus of hotel reviews, it can learn the patterns and structures of typical reviews, enabling it to generate new reviews that are coherent and contextually relevant.

3.3. BERT

To achieve the best rendering both in the proceedings and from the CD-ROM, we strongly encourage you to use Times-Roman font. In addition, this will give the proceedings a more uniform look. Use a font that is no smaller than nine point type throughout the paper, including figure captions.

3.4. TEXT GENERATION

web scraper is utilized to gather data from TripAdvisor, specifically focusing on reviews from a single hotel, which served as our dataset for text generation. In data preprocessing, the n-gram word is used to tokenization the dataset and make the next word a predicted label in the generated text. In this study, the LSTM model is applied to do text generation because the LSTM model can capture long-range dependencies in sequential data very well, which provides a brilliant choice for text generation tasks. In order to enhance the quality and coherence of text generation, the LSTM model is utilized with four layers including the embedding layer, LSTM layer, Dropout layer, and Dense layer to prevent overfitting and ensure robust generalization. In the end, perplexity and accuracy are achieved to evaluate the training model for text generation.

3.5. TEXT SUMMERIZATION

Initially, by set the two empty lists are initialized to record word counts for text and summary sections. Subsequently, the code iterates over the text and summary columns of the DataFrame, calculating and appending word counts to the corresponding lists. It also ensures that if the summary column contains lists, the elements are joined into single strings. Following this, a new DataFrame is created to store the word counts for text and summary sections. Finally, a histogram is generated from this DataFrame, displaying the distribution of word counts for both text and summary data.

Then, by prepare text data for a sequence-to-sequence model. It splits the data into training and validation sets, converts text data into lists of strings, and tokenizes the target summaries. The tokenizer converts text sequences into integer sequences, which are then padded to ensure uniform length. Finally, the vocabulary size for the target summaries is calculated based on the word index from the tokenizer.

Then, there outlines the construction of a sophisticated sequence-to-sequence model with attention, designed for tasks like machine translation or text summarization, which can be applicable to hotel review datasets. The model architecture is composed of an encoder and a decoder, both utilizing Long Short-Term Memory (LSTM) layers.

Then, it came to encoder processes the tokenized input sequences to generate hidden states, which are crucial for the decoding process. The encoder is comprised of multiple LSTM layers, allowing it to capture intricate patterns in the input data. Each LSTM layer returns sequences, as well as the final states, which are then fed into the subsequent layers. Subsequently, the decoder is set up to generate the output sequences (y_train). Similar to the encoder, the decoder employs an LSTM layer. However, in this case, the LSTM layer is initialized with the final states from the encoder, enabling the decoder to generate sequences based on the learned representations from the input data. An attention layer is introduced to enhance the model's ability to focus on different parts of the input sequences while generating the output. This attention mechanism aids in improving the overall performance of the model, particularly in tasks where long input sequences need to be summarized or translated accurately.

After finish previous step, the model setup and implements functions for decoding sequences, converting sequences to summaries, and printing the original and predicted summaries for validation data. The encoder and decoder models are defined based on the trained components of the main model. The decode_sequence function is used to generate summaries for input sequences using the trained model. It initializes the target sequence with the 'start' token and iteratively predicts the next word until the 'end' token or the maximum summary length is reached. Finally, the code iterates over a subset of the validation data, tokenizes the text, pads the sequences and summaries for comparison.

4. EXPERIMENTATION AND RESULT

4.1. TEXT CLASSIFICATION

The activation functions used in the LSTM and BERT models are 'softmax' for text classification, with 'Adam' as the optimizer. When compiling the models, the chosen loss function is 'categorical_crossentropy'. The following is the summary of the LSTM model and BERT model.

```
Layer (type)                 Output Shape              Param #
================================================================
embedding (Embedding)        (None, 500, 100)          729500

lstm (LSTM)                  (None, 128)               117248

dense (Dense)                (None, 2)                 258

================================================================
Total params: 847006 (3.23 MB)
Trainable params: 847006 (3.23 MB)
Non-trainable params: 0 (0.00 Byte)
```

Fig.1. LSTM model in text classification

Layer (type)	Output Shape	Param #
bert.embeddings.word_embeddings	(None, 512, 768)	23440896
bert.embeddings.position_embeddings	(None, 512, 768)	393216
bert.embeddings.token_type_embeddings	(None, 512, 768)	1536
bert.embeddings.LayerNorm	(None, 512, 768)	1536
bert.embeddings.dropout	(None, 512, 768)	0
bert.encoder.layer.0.attention.self.query	(None, 512, 768)	590592
... (omitting other layers for brevity)		
bert.pooler.dense	(None, 768)	590592
dropout	(None, 768)	0
relu	(None, 768)	0
fc1	(None, 512)	393728
fc2	(None, 2)	1026
softmax	(None, 2)	0

Total params: 30,821,122 (117.27 MB)

Trainable params: 30,821,122 (117.27 MB)

Non-trainable params: 0 (0.00 Byte)

Fig.2. BERT model in text classification

The BERT and LSTM models employ similar components and techniques for text classification. However, there are differences in their architecture and specific implementations.

	LSTM model	BERT model
Accuracy	0.89	0.60
Precision	0.88	0.54
Recall	0.89	0.60
F1-score	0.89	0.54

Fig.3. Accuracy comparison of a table in text classification

The experiment expects that the BERT model will perform better than the LSTM model. However, the table indicates that the LSTM model has superior performance over the BERT model in the text classification.

4.2. TEXT GENERATION

Based on the result of text classification, it has been determined that the LSTM model demonstrates superior predictive capabilities. As a result, the decision has been made to employ the LSTM model for text generation purposes. By leveraging the strengths of the LSTM, we aim to enhance the quality and coherence of generated text in one hotel review, especially, descriptions of facilities offered by a particular hotel or detailed insights into their services.

The activation function used in the LSTM is 'softmax' for text generation, with 'Adam' as the optimizer. When compiling the model, the chosen loss function is 'categorical_crossentropy'. The following is the summary of the LSTM model.

```
Layer (type)                 Output Shape              Param #
=================================================================
embedding_2 (Embedding)      (None, 631, 500)          3999500

lstm_2 (LSTM)                (None, 100)               240400

dropout_2 (Dropout)          (None, 100)               0

dense_2 (Dense)              (None, 7999)              807899

=================================================================
Total params: 5047799 (19.26 MB)
Trainable params: 5047799 (19.26 MB)
Non-trainable params: 0 (0.00 Byte)
```

Fig.4. LSTM model in text generation

LSTM model	Epochs = 1	Epochs = 36	Epochs = 72	Epochs = 108
Accuracy	0.038	0.5864	0.6527	0.7074
Perplexity	1093.33	3.5954	2.7184	2.3308

Fig.5. Evaluation of a table in text generation

The evaluation table presents the LSTM model performance of text generation in different numbers of training epochs. At the first epoch, the LSTM model only has an accuracy of 0.038, indicating relatively poor performance. However, as the number of epochs increased to 36, 72, and 108, the accuracy improved significantly from 0.5864, and 0.6527 to 0.7074. This demonstrates the positive effect of training the model for longer durations, leading to enhanced accuracy in text generation. Moreover, the perplexity scores provide insights into the model's ability to predict the next word in a sequence. A higher perplexity score indicates higher uncertainty and less confidence in the model's predictions. In the experiments, the LSTM model exhibited the perplexity decreased from 1093.33 at the first epoch to 2.3308 at 108 epochs, signifying improved prediction power and enhanced coherence in the generated text.

Text Generation Results from Epoch = 1
The Bristol Paris: 'The Bristol Paris Hotel' **Swimming Pool:** 'Swimming Pool Hotel Hotel Hotel Hotel Hotel Hotel Hotel Hotel Hotel Hotel'
Text Generation Results from Epoch = 36
The Bristol Paris: 'The Bristol Paris One Best Hotel Ever Stayed N Really Good Feel Lke Kng Somethng Whle Stayng Food S Great Servce S Really Good Say Ts One Best Hotel N Pars Sure Room' **Swimming Pool:** 'Swimming Pool Ever Come Across Paris Swim Pool Small Glass Champagne Lovely'
Text Generation Results from Epoch = 72
The Bristol Paris: 'The Bristol Paris Hotel Deserves Star Rating Several Year Say Expectation High End Year Stayed Many Times Paris First Time Stayed Le Bristol Many Year Ago Since Last Summer Stayed Two Night Program' **Swimming Pool:** 'Swimming Pool Area One Best Hotel Paris Room Beautiful View Courtyard Great'
Text Generation Results from Epoch = 108
The Bristol Paris: 'The Bristol Paris Hotel Deserves Star Review Hotel Tripadvisor Quite Give Huge Hotel Fail Ready Warmly Welcomed Warmly Margaux Came Assist Drink Bar Escorted U Dark Table Beautifully Arrived Throughout Thing Really Make' **Swimming Pool:** 'Swimming Pool Ever Stayed Recently Night One Best Hotel Ever Stayed Junior'

Fig.6. text generation in different Epoch results.

The text generation in different epochs of train results displays the LSTM model's performance. At the first epoch, the LSTM model appears to continuously repeat words in the text-generated review indicating seriously poor performance. However, as the number of training epochs increased the text-generated review significantly achieved coherent and contextually.

4.3. TEXT SUMMARIZATION

There have text histogram and a title histogram. The text histogram has a peak around 0-50 on the x-axis suggesting that the majority of the text entries in your dataset are quite short, with lengths ranging up to 50 characters or words (depending on the unit of measurement). The frequency quickly drops off for longer texts, indicating that there are fewer long text entries in your dataset. The title histogram has the prominent peak around 5-10 on the x-axis indicates that most titles are within this length range. The smaller peaks near zero and between 10-15 suggest that there are some very short titles and a moderate number of slightly longer titles, but these are less common than titles in the 5-10 range. Then, by continue choose max_len_text equal 100 and max_len_summary equal 4 to get the optimal for model training and text summarization.

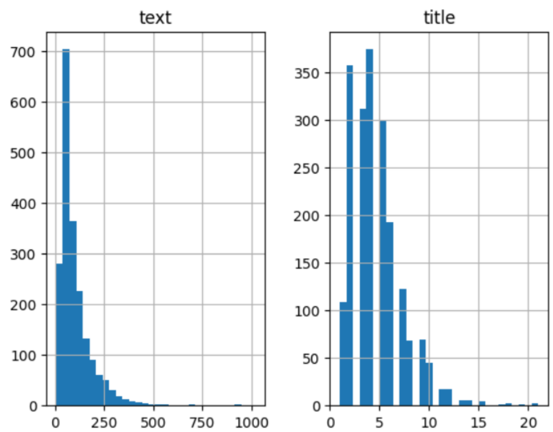

Fig.7. text and title histogram

Then train the model, we train the model several time by using LSTM, this might improve the accuracy of the model. Besides, by use the attention layer, this layer helps the model to focus on specific parts of the input sequence when making predictions, which is especially useful in tasks that require understanding the context, such as machine translation.

Layer (type)	Output Shape	Param #	Connected to
input_layer_60 (InputLayer)	(None, 100)	0	-
embedding_24 (Embedding)	(None, 100, 500)	3,742,000	input_layer_60[0][0]
lstm_48 (LSTM)	[(None, 100, 500), (None, 500), (None, 500)]	2,002,000	embedding_24[0][0]
input_layer_61 (InputLayer)	(None, None)	0	-
lstm_49 (LSTM)	[(None, 100, 500), (None, 500), (None, 500)]	2,002,000	lstm_48[0][0]
embedding_25 (Embedding)	(None, None, 500)	886,000	input_layer_61[0][0]
lstm_50 (LSTM)	[(None, 100, 500), (None, 500), (None, 500)]	2,002,000	lstm_49[0][0]
lstm_51 (LSTM)	[(None, None, 500), (None, 500), (None, 500)]	2,002,000	embedding_25[0][0], lstm_50[0][1], lstm_50[0][2]
attention_layer (AttentionLayer)	[(None, None, 500), (None, None, 100)]	0	lstm_50[0][0], lstm_51[0][0]
concat_layer (Concatenate)	(None, None, 1000)	0	lstm_51[0][0], attention_layer[0][0]
dense_12 (Dense)	(None, None, 1772)	1,773,772	concat_layer[0][0]

Fig.8. LSTM model in text summarization

Finally, there get the summarization from the hotel reviews.

Review: booked room for 13 guest for mountain wed ding celebration from the first reservation call throug h check out all of u felt like royalty yes the decor is b it dated who care the hot tub and pool were an unexp ected go to spot for gathering and relaxing the breakf ast wa always fresh and yummy staff wa attentive did present some challenge that they jumped to respond the location for exploring aspen would be hard to bea t it wa wonderfully memorable weekend facilitated b y the staff and setting at molly gibson the price is not of the quality we experienced we will definitely be b ack
Predicted summary: 'outstanding mountain' day
Review: was extremely impressed with this lodge for the price we were on the lower level and the floor are heated you would expect that for budget accomodations you can even pick which room you want to stay in on their website right outside door is the bus pick up for snowmass so lodge is at an ideal location right in aspen large room with fridge and very clean had place right inside door of room for ski and board perfect set up highly recommend we will stay again
Predicted summary: i've i've memorable

Fig.9. the results in text summarization

Based on predicted summarization, some predictions are fitted with the original review, for the first one, we can see there have “mountain” word in the original text, and the prediction summary has a similar output to summarize the outstanding mountain view. The second review, the “impress” in the sentence, then the prediction gives the “memorable”, which is very similar mean with the original text.

5. DISCUSSION

The performance of the BEAT model may be impacted by the diversity in dataset labels, which is a substantial challenge in text classification. This disparity is most noticeable in datasets that represent consumer reviews of hotels, where the majority of ratings are between four and five stars. As a result of the data imbalance, reaching optimal performance becomes intrinsically challenging. Moreover, the text generation training procedure requires a substantial time investment. In a limited time, 108 training epochs are only able to be completed. Despite this, the achieved performance is still respectable, with a reported accuracy of 0.7 and perplexity of 2.33. However, requires further training and adjustment to reach higher performance levels. Parallel to this, challenges in text summarization include the inability to load necessary packages for attention layers, which forces improvisation. Additionally, the lack of data further jeopardizes model accuracy, highlighting the significance of creative solutions and reliable data pipelines for enhanced performance.

6. CONCLUSION

In conclusion, our research has highlighted the importance of sentiment analysis in improving customer service and business strategies in the hotel industry. By scrutinizing and interpreting large volumes of reviews, researchers and industry experts can uncover trends, patterns, and insights that inform strategic decision-making and drive ongoing improvement.

Despite the challenges and limitations of sentiment analysis, such as the nuances of language and the impact of fake or biased reviews, our study has shown that it can provide valuable insights for both travelers and hoteliers. Future research in this area could focus on refining sentiment analysis models, exploring new data sources, and considering the ethical implications of analyzing customer feedback.

Overall, sentiment analysis offers a powerful tool for enhancing the guest experience and driving business success in the competitive hospitality industry. It is our hope that this thesis contributes to the broader understanding of sentiment analysis in hotel reviews and inspires further research in this important field.

7. REFERENCES

[1] Bulchand-Gidumal, J., Melián-González, S., and González Lopez-Valcarcel, B. (2013). A social media analysis of the contribution of destinations to client satisfaction with hotels. Int. J. Hospit. Manage. 35, 44–47. doi: 10.1016/j.ijhm.2013.05.003

[2] Chu, M., & Chen, Y. (2022, September 28). Language interpretation in travel guidance platform: Text mining and sentiment analysis of TripAdvisor reviews. Frontiers. https://doi.org/10.3389/fpsyg.2022.1029945

[3] Martín C. A., Torres, J. M., Aguilar, R. M., & Diaz, S. (2018, October 23). Using deep learning to predict sentiments: Case study in tourism. Complexity. https://doi.org/10.1155/2018/7408431

[4] Havrlant, L., and Kreinovich, V. (2017). A simple probabilistic explanation of term frequency-inverse document frequency (TF-IDF) heuristic (and variations motivated by this explanation). Int. J. Gen. Syst. 46, 27–36. doi: 10.1080/03081079.2017.1291635