How to raise your rating on Yelp - for bubble tea shop

Jiaying Jia, Qilu Zhou, Zhengyuan Wen

1 Introduction

Bubble tea is becoming more and more popular among the younger generation. Our group aims to explore how to raise the rating for bubble tea shop. The dataset we used is from Yelp related to bubble tea shops. By analyzing attributes and reviews of different shops, we developed intuitive and straightforward suggestions to bubble tea shops.

2 Data Cleaning

We selected bubble tea shops from the whole Yelp data set by filtering for shops which include "bubble tea" in their category.

After getting the data of bubble tea shops, we deleted 10 shops whose business id was missing. And when we analyzed the attribute "WiFi", there were 2 shops offering paid WiFi but significantly affected stars, so we deleted these two shops. Moreover, we calculated the total open hours of each shop for a week based on their open hours from Monday to Sunday and deleted those whose open hour were less than 15 hours per week.

Finally, 686 shops remained. The data includes variables in four aspects, that is, basic information about the shop (business id, name, state, city etc.); all the 34 attributes; open hours (daily open hours, weekly open hours, open hour level); key word information from reviews (sentimental scores, mean values, standard deviation).

3 Business Attributes Analysis

The original data contains 34 attributes and we wanted to find which attributes would have significant influence on star rating, so we conducted an ANOVA on star rating to all the attributes. Since we have 686 shops, we can assume the data is normal on the response. And because the data for each shop is independent, the independence assumption is also satisfied.

The ANOVA results show that WiFi, weekly open hours, Outdoor Seating, Wheelchair Accessible, Noise Level are significant. For all significant attributes, we wanted to find out how each level of the attributes affect stars, so then we ran a linear regression. The results of significant attributes levels in the regression are shown in Table 1. When shops offer free WiFi, their stars would increase by 0.39; when they offer outdoor seating option, their stars would increase by 0.29; when the noise level is quiet, the stars would increase by 0.20, while it would decrease by 0.84 if the shops are very loud. We divided weekly open hours into three levels: less than 8 hours a day ("< 56"), 8-12 hours a day

("56-84") and more than 12 hours a day ("$>$ 84"). We can see from the result that for bubble tea shops, contrary to popular belief, shorter open hours brings a higher rating. We also compared average stars to different levels of attributes. We found that if open time is less than 56 hours a week, the average star rating would be 0.4 higher than those open more than 84 hours a week; shops with free WiFi would be 0.2 stars higher than those who don't; for shops offering outdoor seating, their rating will be 0.2 higher than those who don't; and for the environment, quiet shops tend to get 1 star rating higher than noisy shops.

Table 1: Regression Estimate of Significant Attributes

Variable	Estimate	Std. Error	t value	P value
Intercept	3.90	0.17	23.09	<2e-16
Weekhour 56-84	-0.09	0.05	-1.83	0.07
Weekhour > 84	-0.31	0.10	-3.10	0.002
WiFi no	-0.16	0.06	-2.63	0.009
WiFi none	-0.23	0.07	-3.59	0.003
BusinessAcceptCreditCards	3.12	1.56	5.24	0.006
OutdoorSeating None	0.20	0.05	3.68	0.0003
OutdoorSeating True	0.10	0.06	1.76	0.0798
NoiseLevel quiet	0.20	0.07	2.76	0.006
NoiseLevel very_loud	-0.85	0.39	-2.19	0.03

The above results indicate that people who go to bubble tea shops might not just grab one and leave. They probably also have the need to sit down to study or read in the shop so free WiFi is a necessity. Shorter weekly open hours probably guarantees the quality of raw materials and ingredients as well as the freshness of tea. Shops with wheelchair accessibility may tend to have larger space, which is positively related to star rating. And people prefer quieter environment when they need to work or study in the shop, so noise level is also important.

4 Review Analysis

To analyze reviews, we developed a model to capture the "sentiment" by outputting the rating for specific sentences in the reviews. The output of rating ranged from 1 to 5. The higher, the better that the comment is.

4.1 How to obtain a sentiment score?

We used a pre-trained model called BERT [1] (Bidirectional Encoder Representations from Transformers [2]). It's a state-of-art technique used in processing natural language. We did a supervised learning via a neural network based on the smallest BERT. The output is the rating of a review, the input is a review. Here is the structure of our neural network. See Figure 1.

For training process, we split the dataset in a ratio of 2:8 for validation and training. The batch size is 32. We used Adamw optimizer with 10% warm up steps, set the loss function to be MSE. The initial learning rate is 3×10^{-4}. After training for 4 epochs, we found the loss of validation set tended to be stable. See Table 2.

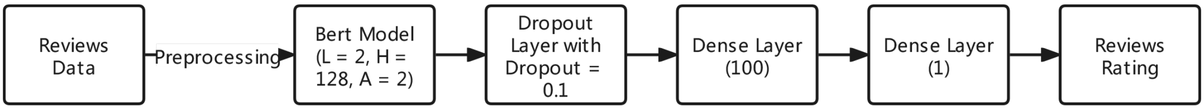

Figure 1: Structure of model

Table 2: Training and validation loss (MSE)

Epochs	1	2	3	4
Training Loss	0.7899	0.4795	0.3810	0.3415
Validation Loss	0.5445	0.5255	0.4958	0.4957

Here are some examples of how the model works. Given some sentences, *Great selection of bubble tea. Taste good. Just a bit slow on service. When we came place is packed and it's 9:15pm. Not enough space to sit.* The model will give a predicted value 3.4802856. *This is my FAVORITE place to get bubble tea.* The model will give a predicted value 4.9756384. The actual rating of the above sentence is 3, 5, which are similar to the predicted values.

The main purpose of proposing this model is to obtain a sentiment score not only for a review, but also for a sentence. Just like the first example mentioned above, the overall rating of review is 3, but the reviewer actually likes the bubble tea taste and price, it's the service speed that drags the rating down. So, using our model, we could give *Great selection of bubble tea* a 4.592097 rating, and *Just a bit slow on service.* a 2.5831966 rating, which is consistent with human intuition. We could give a reasonable rating for key words in each sentence of our model.

4.2 Analysis with sentiment score

4.2.1 Six overall aspects

We defined 6 aspects to describe the overall performance (rating, score) of a bubble tea shop. Each aspects have different criteria. Followings are the aspects corresponding to criteria. To better illustrate the results, avoiding the negative and positive effect brought by the word itself, we only used neutral nouns in the analysis. See table 3.

Table 3: Words used to calculate ratings of different aspects

Aspects	Atmosphere	Price	Service	Food
Word(s)	atmosphere, ambiance, environment	price	service	food, drink
Aspects	Tea Ingredients		Tea Types	
Word(s)	tora, bubble, boba, herbal, rainbow		black/oolong/green tea, matcha, jasmine	

For each sentence containing the above key words, we predicted an estimated rating. Then we took the mean for each word as a representative sentiment score of a key word. At last, we computed a weighted average (weight is sample sentences) from the key words in each aspect as the sentiment scores of an aspect. As a result, we could obtain the scores of six aspects for different shops. At last, the average of six aspects was exactly the overall score. Here are the scores for all 686 bubble tea shops. See Table 4.

Table 4: Sentiment Scores (Ratings) for all 686 bubble tea shops

Aspects	Atmosphere	Price	Service	Food	Ingredients	Tea Types	Overall
Scores	4.213204	3.376005	3.704744	3.460416	3.777528	3.888872	3.736795

The scores in Table 4 ranges from 1 to 5. The predicted values which are larger than 5 or less than 1 will be adjusted to 5 or 1. Basically, most of the number are within the range 1 to 5, and even if it was higher than 5 or lower than 1, it only exceeded around 0.05 in most of the cases. A finding from sentiment scores for all bubble tea shops is that customers complained about the price a lot, and most of the bubble tea shops did great in atmosphere.

4.2.2 Suggestions and strengths for a specific bubble tea shop

We further extracted some information based on key words. Some examples of key words are green tea, oolong tea, black tea, matcha, etc. We calculated the sentiment scores for each key word, and compared them with the overall level of that key word. We used Welch's t test to test whether the difference between a specific shop and overall level is significant at 5% significant level. If it did, then we would consider putting that word into our suggestions or strengths for the bubble tea shop since the results suggest that it would do better or worse than the overall level. A finding from the overall level based on key words is that oolong tea (4.142) is the most delicious compared with other tea types like green tea (4.104), black tea (4.015), matcha (3.864) and jasmine (3.838) by Welch's t test.

5 Shiny App Design

We developed a shiny app that allowed users to find bubble tea shops by choosing state, city and shop name. The app contains three pages: Basic Information, Rating Analysis and Performance. In the section of Basic Information, we provided basic information about the shop including its star rating, number of reviews, detailed address and open hours. A bar plot can also be seen in the page showing the star rating from six aspects (that is "Atmosphere", "Food", "Price", "Service","Ingredients" and "TeaTypes") with a red line showing the mean value of these six aspects and a blue line showing the overall performance of all shops. In the Rating Analysis page, we drew four radar plots comparing this shop with all shops from 4 aspects: overall, ingredients, tea types and dessert. In the Performance section, we gave customized strengths and suggestions.

6 Strengths and Weaknesses

Our results are very clear and easy to interpret with detailed, thorough and customized suggestions from tea types, ingredients to ambiance and Wifi. Following the suggestion, shop owners can accurately improve their performance.

However, the algorithm we used would run for a long time. We used sentence sentiment scores (might contain several words) to substitute word sentiment scores, which might cause bias. Our ANOVA and regression analysis also produced some results that are hard to interpret, for example none accept business credit cards have the highest stars.

References

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017.

Contributions

1) QZ (Qilu Zhou) was mainly responsible for business attributes analysis, wrote the related code such as ANOVA analysis on business attributes and data cleaning based on missing value in Business Attributes. QZ also wrote the first three sections (Introduction, Data Cleaning and Business Attributes Analysis), had partial contribution on Shiny app and the last section in this summary report. QZ completed the most of the work for presentation slides based on our summary report.

2) ZW (Zhengyuan Wen) was mainly responsible for reviews analysis, built the Bert model to obtain sentiment scores for each sentence, wrote the related codes for extracting the bubble tea stores/reviews, reviews analysis, writing the suggestions and strengths for specific bubble tea shops. ZW also wrote the fourth part Review Analysis in this summary, and had a little contribution on presentation slides based on the version of QZ provided.

3) JJ (Jiaying Jia) was mainly in charge of Shiny App Design, wrote all the codes related to shiny app design such as bar plot, UI design and back end implementation. JJ also did some adjustment on the final data for easier interpretation on Shiny App. JJ wrote the fifth part Shiny App Design in our summary report, and was responsible for polishing the whole summary report.