

Analysis of Pizza Restaurants

STAT 628 Module 3 Group 2 :

1. Introduction&Background

Pizza Chains almost spread in every city of North America. Many people are willing to choose pizza as their first choice when they have no idea what to eat. In Module3, we decided to assume Pizza Hut as our object to have analysis. And the other 9 pizza chain restaurants, which are selected by the quantity of the restaurants across the nation, are also analyzed and compared with Pizza Hut. For operators of Pizza Hut, it's important for them to get a full understanding about themselves and their competitors, their own or their competitors' advantages and disadvantages to improve themselves.

With the given data *user.json*, *review.json*, *tip.json* and *business.json*, all of these four datasets contain many variables and observations need to be filtered. So we only kept the observations related to 10 pizza chain brands which have the largest amount of restaurants, defined as "top 10 pizza chain".

Here comes our goal of this thesis and our work for it:

- During the process of analysis, users' activity and influential levels are taken into consideration.
- the method:
 - (a) **local weighted linear regression, logistic regression and LDA model** when selecting the important adjoined words that mostly influence the stars of restaurants
 - (b) employ **user weight as adjustment** when constructing feature matrix and y labels(modified stars)
 - (c) find the noun words adjacent to the selected adjoined words to form the **noun-adj pairs**.
- Find out the strongness and weakness of Pizza Hut using SWOT table. And compared with other pizza chain brands to gain final advices.

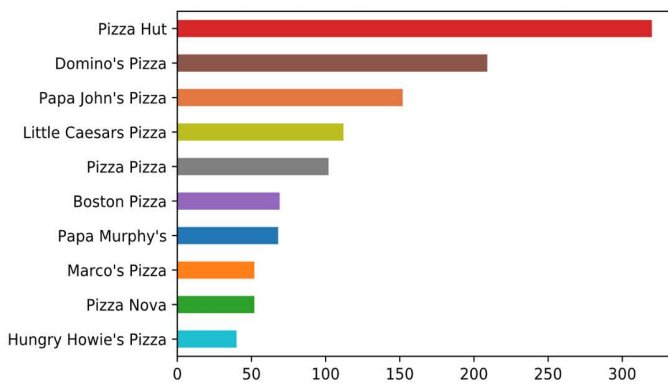
2.Data Description

2.1Top 10 Chain Pizza Restaurants

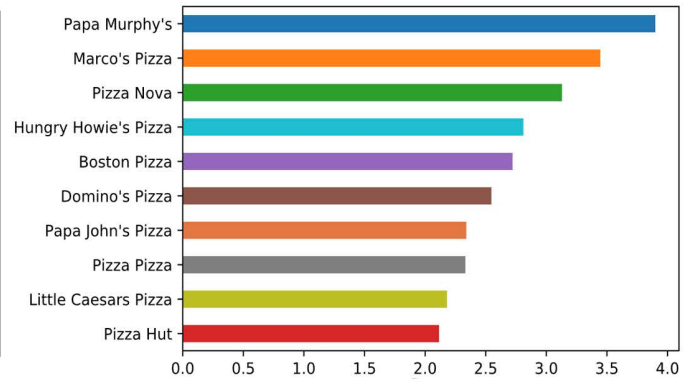
Introduction

By scanning through the business.json file, we could identify different kinds of business in the dataset, including finance institutions, restaurants, hair and beauty and so on. Firstly we focused on the restaurant category and there are still many kinds of restaurants such as Chinese restaurants, Japanese restaurants and fast food restaurants. Then zoom in the fast food restaurants and select only the pizza restaurants, we could then count the number of these pizza restaurants in the dataset. Because many pizza brands only have one or several restaurants, we sorted them by the quantity of the restaurants and chose only the top 10,finally we got the object we will study on.

Then, we did some basic statistical analysis with these top 10 pizza restaurants. We got the locations of them and identify that based on our dataset, the restaurants are located in 11 states **AB, AZ, IL, NC, NV, OH, ON, PA, QC, SC, WI**. Then we first compared the average stars of the top 10 pizza restaurants. Second, we drew the average star for Pizza Hut restaurant to get a overall review.



Plot 1: Top 10 Pizza restaurants count



Plot 2: Top 10 Pizza restaurants average star

From plot1 and plot2 we know, Pizza Hut has the lowest average star although it has the most stores. Hence we do further analysis with Pizza Hut.

2.2 User Weight

Introduction

Considered about the fact Yelp is not only a website/mobile app which could show you different business directory and public assessments. To some extent, it is also a social media and a review forum where users are able to communicate with each other. So, it will be likely to think that maybe we tend to believe special classes of users more than others.

Let's assume that you are hungry and want to adopt a restaurant. There are only two users' reviews but you are only able to pick one of them to help you make the decision. The first user has lots of followers and writes many reviews. When you look at his/her files, a large number of different kinds of compliments received by the user. But another one, just like an almost blank account. Which user do you tend to choose?

On social media, there is a common phenomenon that people are more likely to choose a restaurant or a product recommended by influential users or active ones. Naturally, we tried to measure users' influence and activeness on the social network. When we are in the face of social network analysis, centrality is one of the concepts studied most. A wealth of methods have found centrality measures to identify special users. Between millions of users, how to measure new definitions we will use later, in other words, how to work with the data to calculate users' weight. Companies like Twitter have diverse measures.

At first, PageRank algorithm came into our views. PageRank algorithm is an effective method which is used to rank websites and there are several measures based on it. But when coming back to browse our users dataset, we could find that this dataset doesn't include user followers' specific information so that it's almost impossible for us to build the network of fellowship between different users.

Entropy Weight Method

On account of the dataset we only contain finite variables, the entropy weight method is the final choice. In order to include more aspects, review counts, amount of followers, number of useful/funny/cool votes sent by the user, and number of plain/funny/cool compliments received by the user, these measures of activity and influence are able to gain users' weight.

The procedure of entropy can be expressed in six steps [1]:

- Normalized the raw data to eliminate anomalies with different units and scales. Because later we will use logarithm to process data, a special kind of normalization is used here:

$$x'_{ij} = \frac{x_{ij} - \min(x_{1j} \dots x_{nj})}{\max(x_{1j} \dots x_{nj}) - \min(x_{1j} \dots x_{nj})}$$

- Compute the porproction p_{ij} :

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}}, i = 1, 2, \dots, 15442, j = 1, 2, \dots, 8$$

- Compute the entropy e_j of arribute j:

$$e_j = -\frac{1}{\ln(n)} \sum_{i=1}^n p_{ij} \ln(p_{ij}), j = 1, 2, \dots, 8, p_{ij} \ln(p_{ij}) \text{ is defined as 0 if } p_{ij} = 0$$

- Compute d_j as the degree of diversification.

$$d_j = 1 - e_j, j = 1, 2, \dots, 8$$

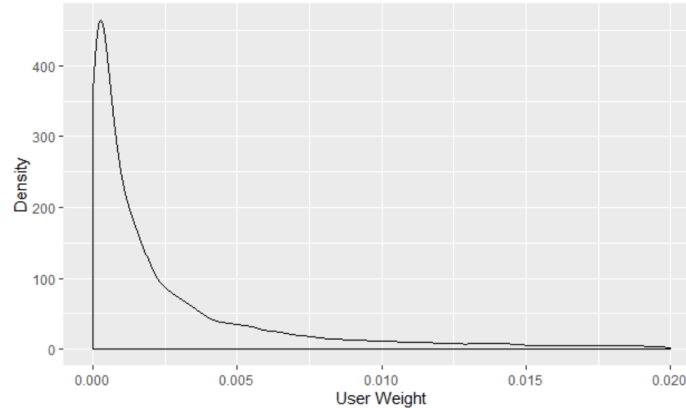
- Compute w_j as the degree of importance of attribute j.

$$w_j = \frac{d_j}{\sum_{j=1}^8 d_j}, j = 1, 2, \dots, 8$$

- Finally, compute the score(weight) s_i of every user.

$$s_i = \sum_{j=1}^8 w_j x_{ij}, i = 1, 2, \dots, 15442$$

After computing user weights, the density plot is shown in Plot3. But on account of most data cumulating on the left and close to 0 we only chose [0 0 02] to demonstrate



Plot3: Density plot of user weight(from 0 to 0.02)

2.3 Important adjected words in Reviews

Review data cleaning

After selecting the reviews from the top 10 pizza restaurants, altogether 17951 reviews, the first procedure is to match these reviews with the restaurants' name and user weight we computed before. Then we focused on the review text. To begin with, we did translation. It could be identified that there are more than 10 countries' languages (**Afrikaans,Danish,French,...Spanish**). By using the package in python,we translated them all in English.

Secondly, we changed each review into sentences and deleted those sentences contain 'if' in them, because these conditional sentences are usually virtual and cannot correctly present the current situation of the restaurant. These 'if' sentences are the first direct suggestions we could get from the customers. Next, by removing positive helping verbs, changing sentence pattern ('**not only, but also**' -- '**...and ...**'), extending negative abbreviation('don't--do not'), combining adverbs with words behind them ('**not great**'-- '**notgreat**'),deleting punctuation marks, deleting positive helping verbs again and combining negative patterns again further ('**not great pizza**' -- '**notgreatpizza**'), we condensed the review text.

Extract adjected vocabulary and form the feature matrix

Importing **nltk** package gives adjective labels 'JJ'(adjective words),'JJR' (comparative),'JJS'(superlative),'RB'(adv),'RBR','RBS' [2] as objective tags and then selected the words that labeled by these tags. We could get a list of adjectives(adv). Then after counting these words throughout the reviews and sorting them, we chose first 50 words with higher frequency and formed a list called adjective vocabulary list. Similarly we formed a vocabulary list that contained 50 tokenized negative phrases ('**neveragain**','**notgreatpizza**') as not vocabulary list. Combining these two lists and we got a vocabulary with 100 words. Then, we could construct a 17951* 100 matrix **X**, the ijth entry represents the frequency of the jth word in the ith review. The **y** labels are just the star of each review.

3. Model construction

Local Weighted Linear Regression(LWLR)

- The coefficients of LWLR could be computed by:

$$\hat{\beta} = (X'WX)^{-1}(X'WY)$$

where W is the weighted of each review sample, that is to say W[i,i] represents the ith user weight. The larger the userweight, the more important the sample it counts in the model fitting.

To construct the model first we split the feature matrix **X** and the label **y** into training set and test set by ratio 8:2. We did this because we'd like to apply the LWLR model and the original linear regression model to the test set and compared the MSE of them to identify the model accuracy. After sorting the coefficients from largest to smallest, we displayed a few coefficients that are are smallest 4 for these 2 models.

notready	-0.2841758047183900	notgoingback	-0.2975867135047550
neverordered	-0.3023918838235440	neverordered	-0.3013204369443360
later	-0.3279864851082150	later	-0.3303707270438090
neverreally	-0.6195962103424990	neverreally	-0.7563027667217800

Table 1: The weight for each adjective words (left is without user weight adjustment)

It could be identified that after using the user weight as adjustment, the LWLR model has different rank for the adjectives. That is to say user weight has influence in the final decision making.

To compare the 2 models, we could get the MSE of these two models. The result is shown in Table 2. Hence from the results we find that after using the user weights as adjustment, the MSE of the LWLR model is smaller than that of the original linear regression model. Although there might be some randomness in this result since the train and test set might be split and chosen in a different way, the result can still give us some promising picture that the application of the user weight somewhat gives positive influence to the model construction.

LDA Model with Local Weighted Tf-Idf

- Local weighted Tf-Idf: For the tf-idf term of the i -th word in the j -th document constructed by the j -th user

$$widf_{i,j} = w_j * lg(\frac{|D|}{|\{j : t_i \in d_j\}|})$$

where w_i is the weight corresponding to the j -th user, D is the total number of all document and the denominator is the number the document which includes the i -th word.

After that we used the output of modified LDA model as the input of the regression or classification model. In this step, we chose linear regression, binary logistic regression and random forest to exam the performance of this composite model.

Model	OLR	LWLR	LDA	LWLDA
MSE	3.933	3.683	1.990	1.934

Table2: MSE of 3 models.(OLR:Original Linear Regression.LWLR:Local Weighted Linear Regression.LDA:Latent Dirichlet Allocation.LWLDA:Local Weighted Latent Dirichlet Allocation)

Based on Table2, we could find after user weight adjustment, the MSE of linear regression model and LDA both decrease, indicating improvement of the models. Consequently, we chose LWLDA model because it has smallest MSE to select adjectives.

Then we detected adjective-noun pairs in the original review file and calculated the quantity of positive and negative pairs. From these pairs information we could derive the average altitude of customers in a specific perspective. Then we could test whether there is significant different between different restaurants from the perspective defined by the noun.

The sentiment table of noun word "pizza"		
	Pizza Hut	Other 9 Pizza Chain
Positive	74	322
Negative	108	200

Table3: The sentiment table of noun word 'pizza'

After applying chi-square test, we can certainly admit that there is a significant difference between Pizza Hut from other 9 pizza chain since the p-value is 1.318e-06

4. Final suggestions

SWOT Analysis

Based on what we analyzed above, the final task is to give advice. SWOT analysis is used here. According to Wikipedia and combining with our case, SWOT is a pretty useful strategic planning technique which should help owners of Pizza Hut identify strengths, weaknesses, opportunities, and threats related to business competition.

When we compared the keywords from Yelp reviews between Pizza Hut and other 9 pizza chain. In most aspects, the difference between them is not that great. But strength and weakness could still be compared below:

The noun and "Fast" pairs(in every 1278 nouns)			The noun and "Delicious" pairs(in every 1278 nouns)		
Noun	PizzaHut	Other 9 Pizza Chain	Noun	PizzaHut	Other 9 Pizza Chain
Delivery	39	31	Pizza	15	35
Service	32	17	Food	5	32
Pizza	7	5	Crust	1	17

Table4: Frequency tables for adjectives near nouns "fast" and "delicious"

And more kinds of nouns appear in other 9 Pizza Chains, like gralic, dessert, pepperoni, cream, jalapeño, ravioli, breadstick, sangria, cheesecake, burger. It might be something we could improve.

Most pizza chains have similar keywords in reviews. Although Pizza Hut has just got the lowest average star in these ten pizza chains, there is still the largest amount of Pizza Hut restaurants across our dataset. If operators of Pizza Hut decide to roll their plans out, they will be able to afford these when other chains might struggle with the capital.

But on the other hand, due to the amount of Pizza Hut chains, it's might be a challenge for them to make the decision. With great revenue comes great operating risk. Conservative measures might be more likely to be chosen when the decision-makers are not firm enough to diffuse new strategies.

Based on what we have analyzed above, here is the final SWOT table.

SWOT Analysis	
Strength	High work efficiency, delivery on time
Weakness	Poor selection menu, not so tasty
Opportunities	Brand impact, well-funded to carry out business plan
Threats	Hard to diffuse new technology or new strategy to all chains

5. *Strength and weakness*

Strength:

- Consider highlighted users when constructing models to extract the adjectives for our final suggestions making.
- An accurate method to detect the sentiment of customers based on their reviews

Weakness:

- When using LDA model, we assume words are exchangeable, and don't consider the sentence structure.
- We might have not mine enough information from short reviews

6. *Contribution*

- Hao Qin: LDA, regression and classification model, data analysis
- Jiacheng Miao: Shiny Application construction.
- Shirley Zhang: Review data cleaning, Linear Regression, Local weighted linear regression constructed.
- Xiyue Wang: Compute user weight, SWOT analysis, write summary

Reference

[1] Imprecise Shannon's Entropy and Multi Attribute Decision Making:<https://ui.adsabs.harvard.edu/abs/2010Entrp..12...53L/abstract>
(<https://ui.adsabs.harvard.edu/abs/2010Entrp..12...53L/abstract>)

[2] NLTK property for words and their tags: <https://blog.csdn.net/john159151/article/details/50255101>
(<https://blog.csdn.net/john159151/article/details/50255101>)

Shinny Link

<https://jiacheng-miao.shinyapps.io/Module3/>