A Recommender System for Cultural Restaurants Based on Review Factors and Review Sentiment

Emergent Research Forum (ERF)

Sonya Zhang
California State Polytechnic University, Pomona
xszhang@cpp.edu

Mohammad Salehan
California State Polytechnic University, Pomona
msalehan@cpp.edu

Andrew Leung
California State Polytechnic University, Pomona
acleung@cpp.edu

Ishmene Cabral
California State Polytechnic University, Pomona
ircabral@cpp.edu

Navid Aghakhani
University of Tennessee at Chattanooga
navid-aghakhani@utc.edu

Abstract

Online consumer reviews are becoming a key part of choosing a local business, with more consumers than ever turning to the Internet for help with everyday decisions. These reviews can help increase the visibility of the businesses, as well as provide invaluable business development insights for the owners. However, the vast amount of reviews and limited resources can make it difficult for a business to extract intelligence that helps them decide which area(s) for improvement to focus on. Previous studies have suggested that restaurant customer reviews can be categorized into multi-factors such as service quality, product quality, menu diversity, price and value, atmosphere, etc. Consequently, drawing upon eight restaurant review factors from literature and cultural restaurant reviews from a recent Yelp dataset, we propose and evaluate a content-filtering recommender system that automatically classifies individual reviews, predicts the weight and sentiment of each factor in the review, and summarizes the significant area(s) for improvement for each cultural restaurant category. We expect the findings to vary among different culture categories of restaurants. This recommender system helps to automate mining the ever growing online reviews, and provide specific business development insights for cultural restaurants. It is also potentially for other types of business with some modifications on the review factors.

Keywords

Restaurant reviews, Yelp.com, recommender system, culture, machine learning.

Introduction

Online consumer reviews provide significant value to both consumers and businesses. More and more consumers now read online reviews of a product or service before making a purchase decision (Salehan and Kim 2016, Salehan et al. 2017). In response, businesses are also monitoring and reacting to online customer reviews to obtain customer feedback and suggestions to improve quality of products or services, and ultimately gaining substantial financial performance. (Anderson and Magruder 2012, Luca 2016).

As online review websites such as Yelp and TripAdvisor grow over time, businesses received a large amount of customer reviews on their products / services (Salehan and Kim 2016; Salehan et al. 2017).
Many studies investigated predictors of review usefulness (e.g., Connors et al. 2011; Mousavizadeh et al. 2015; Mudambi and Schuff 2010; Salehan and Kim 2016; Salehan et al. 2017). The large quantity of reviews can make it difficult for businesses to decide which reviews to pay attention to or summarize significant information for actions, while using technologies such as natural language processing and machine learning to discover subtopics from these reviews can be very useful (Salehan and Kim 2016; Salehan et al. 2017).

In the context of restaurant reviews, Yüksel and Yüksel (2002) identified 9 different restaurant characteristics from customer feedback, such as service quality, product quality, menu diversity, price and value, atmosphere, etc. Using these restaurant characteristics/factors, we propose a content-filtering recommender system that automatically classifies each review at sentence-level, and evaluate the factor weight and sentiment for various cultural restaurant categories. Restaurants can use such knowledge to make informed business development decisions by focusing on the identified area(s) of improvement. The remaining part of this paper is organized as follows. First, we will review the related literature. Next, we discuss our methodology and preliminary results. Finally, we conclude with the expected contributions.

### Literature Reviews

#### Restaurant Reviews

Restaurant experiences consist of both tangible products and intangible services components (Bojanic and Rosen 1994). Restaurant service providers not only interact with their customer through price, quality, and variety of food but also through quality and speed of service and dining environment. A customer’s experience with an unpleasant service environment may negatively affect his/her cognitive, emotional, and physiological response, which in turn may negatively influence his/her prospects about people/restaurant staff or product/food there. Therefore, customer review of restaurant experience is a complex process that involves processing a multitude of factors (Yüksel and Yüksel 2002). Identifying dimensions and attributes contributing to customer satisfaction in restaurants can provide practical knowledge for restaurant management to improve their food, service quality and customer satisfaction. Based upon a survey of 449 tourists who dined at independent none-fast-food restaurants, and using factor analysis of customers’ perceived performance ratings, Yüksel and Yüksel (2002) identified 9 restaurant evaluation factors: (1) service quality, (2) product quality, (3) menu diversity, (4) price and value, (5) atmosphere, (6) healthy food, (7) location, (8) smoke, and (9) visibility. We dropped the 8th factor, smoking, because it is banned in indoor public spaces and workplaces in U.S. and Canada.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Service quality</td>
<td>The standards, consistent quality, courtesy of the service; friendliness, knowledge, communication, competency, and attentiveness of the server.</td>
</tr>
<tr>
<td>2. Product quality</td>
<td>The quality, portions, tastiness, temperature, presentation, preparation consistency and non-greasiness of food.</td>
</tr>
<tr>
<td>3. Menu diversity</td>
<td>The menu variety, or availability of menu items, dishes or beverages liked, local dishes, and health food choice.</td>
</tr>
<tr>
<td>4. Price and value</td>
<td>The prices and value of food.</td>
</tr>
<tr>
<td>5. Atmosphere</td>
<td>The atmosphere or ambience in the restaurant.</td>
</tr>
<tr>
<td>6. Healthy</td>
<td>The availability of healthy food and nutritious food.</td>
</tr>
<tr>
<td>7. Location</td>
<td>The location, crowd level, and operating hours of the restaurant.</td>
</tr>
<tr>
<td>8. Visibility</td>
<td>The visibility of food preparation area.</td>
</tr>
</tbody>
</table>

*Table 1. Restaurant Review Factors*
Cultural Restaurant Reviews

Consumers who enjoy ethnic foods have grown in number and ethnic restaurants have become mainstream in the U.S. (Jee 2013; Goss 2017; Po 2007). Consumers are attracted to restaurants offering different cultural cuisines by various factors (Huang et al. 2014; Nakayama 2015; Camillo 2010). For example, good price is found to be the initial attraction to the customers who dined at Chinese restaurants, while good service and food quality, and pleasant environment of the restaurant help to build up customer’s positive experience and retain customers (Huang et al. 2014). Food and atmosphere are considered far more important than the other common aspects for Ramen restaurants in both U.S. and Japan (Nakayama 2015). Taste, simplicity, and the variety of Italian regional cuisines were factors found to influence the success of Italian restaurants in the U.S. (Camillo 2010). In this study, we hope that automatically mining and extracting such business insights based upon large amount of online customer reviews can benefit restaurants in more cultural categories.

Recommender Systems

Recommendation agents (RAs) “are software agents that elicit the interests or preferences of individual users for products, either explicitly or implicitly, and make recommendations accordingly” (Xiao and Benbasat 2007). RAs are extensively used in the context of e-commerce to make recommendations of suited products or service to customers and provide a type of mass customization on the internet (Ansari et al. 2000; Detlor and Arsenault 2002; Grenci and Todd 2002; O’keefe and McEachern 1998). RAs can be categorized in 3 different ways: (1) content filtering vs. collaborative filtering vs. hybrid, (2) compensatory vs. non-compensatory, and (3) feature-based vs. needs-based vs. hybrid (Xiao and Benbasat 2007).

Content-based filtering methods are based on a description of the item and a profile of the user’s preferences. An example is the personal shopping bot at MySimon.com through the Active Buyer’s Guide. Collaborative filtering RAs generate recommendations based on preferences of like-minded people. The hybrid method consists of a combination of the two methods (Ansari et al. 2000; Xiao and Benbasat 2007). Compensatory methods permit tradeoffs between attributes allowing desirable attributes to compensate for less desirable attributes while non-compensatory RAs do not allow tradeoffs between attributes (Knijnenburg et al. 2011). Finally, feature-based RAs provide recommendation by asking questions about product features the user prefers. Needs-based RAs, in contrast, are appropriate when users are not able to identify features but can describe their needs (Xiao and Benbasat 2007).

The Proposed Recommender System

This study develops a content-filtering recommender system for cultural restaurant categories based on the restaurant review factors and sentiment. The recommender system will classify individual reviews at sentence-level, predicts the weight and sentiment of each factor in the review, and summarizes the significant area(s) for improvement for each specific cultural restaurant category. This recommender system can help automate the review classification, extract business intelligence, and drive informed decision on target improvements for existing restaurant businesses and new opportunities for investors.

Research Methodology

The Yelp Data Challenge Round 9 dataset contains 4.1M reviews and 947K tips by 1M users for 144K businesses in UK: Edinburgh, Germany: Karlsruhe, Canada: Montreal and Waterloo, U.S.: Pittsburgh, Charlotte, Urbana-Champaign, Phoenix, Las Vegas, Madison, Cleveland. In this study, we randomly selected 290 reviews of restaurants located in the U.S. and Canada from the Round 9 dataset. We first parsed each review into sentences (total 2552 review sentences). Then two researchers manually identified each review sentence as one or more of the 8 restaurant selection factors listed in Table 1, as well as the sentiment (positive, negative, or neutral). The intercoder reliability is excellent - Scott’s Pi (p), Cohen’s Kappa (k), and Krippendorff’s Alpha (a) are all 0.96. We then used the coded data to train and predict review factors and sentiment by different algorithms, including SVM, decision tree, random forest, and Naïve Bayes, until we find the best classification model. Such model is then used to predict the review factors and sentiment for a larger set of round 9 data, based upon which recommendations for
improvement and opportunities will be made for different restaurant cultural categories, such as Italian, Mexican and Chinese, suggested by the National Restaurant Association as top categories in the U.S. (2015). An example recommendation would be: Italian restaurant customers talked more about Food quality (weights 50%) and Ambience (weights 25%) in their reviews, and they are generally happy in these two aspects, but they are not too happy about Menu options, which still weights 18%. So new and existing restaurants should make sure to provide good quality food and ambience, while extending and customizing their menu options in order to differentiate from competitors.

Results and Discussion

At this preliminary stage, we have tried a variety of algorithms, including SVM, Naïve Bayes, Random Forest, Decision Tree, for all eight review factors. So far our model using Naïve Bayes for the Service factor yields the best performance, and all other algorithms also yields better than the baseline 50% (see Table 2). To create the classification model, we used multiple techniques, including:

- Filter out stop words - use customized dictionary
- Dimension deduction - create example set to cluster relevant words into a coherent and principle topic, e.g., “waiter”, “waitress” and “bartender”; generate n-Grams terms, etc.
- Feature selection – use binary term occurrence vector and perceptual prune method
- Balance data – up-sample minority class
- 20 fold cross validation

In future we will continue improving the classification models, as well as adding more coded data for training. We can then apply the best models to predict for a larger dataset. Eventually, we can calculate and summarize the weight and the sentiment of each review factors to provide recommendations for specific cultural restaurant categories.

<table>
<thead>
<tr>
<th>Service Factor Performance</th>
<th>SVM</th>
<th>Naïve Bayes</th>
<th>Decision Tree</th>
<th>Random Forest</th>
</tr>
</thead>
<tbody>
<tr>
<td>True 1 Pred. 1</td>
<td>64.11%</td>
<td>79.09%</td>
<td>64.79%</td>
<td>59.59%</td>
</tr>
<tr>
<td>True 0 Pred. 0</td>
<td>86.01%</td>
<td>89.28%</td>
<td>96.11%</td>
<td>98.53%</td>
</tr>
<tr>
<td>F-score</td>
<td>63.12%</td>
<td>92.33%</td>
<td>70.24%</td>
<td>70.98%</td>
</tr>
</tbody>
</table>

Table 2. Supervised Classification on Review Factors and Sentiment – Service

Expected Contributions

Many studies have looked into multi-aspects in online review however few have looked how important each aspect is to consumers and restaurants (Nakayama 2015). This study moves beyond describing the predictors of review factors and provides a practical method for automatically classifying reviews and extracting business insights for each cultural restaurant category. This recommender system helps automate data analysis of the ever-growing online reviews, and provide adaptive business development insights for existing and new restaurants, also potentially for other types of business with some alterations on the review factors. The same mechanism can be applied in the case of an individual restaurant, where the restaurant will be able to get real-time knowledge about the focus area(s) and general sentiment from vast amount of their online customer reviews.

REFERENCES


