Agent-Based Modeling of Consumer Choice by Utilizing Crowdsourced Data and Deep Learning

Boyu Wang[1]
Department of Geography, University at Buffalo, NY, USA

Andrew Crooks
Department of Geography, University at Buffalo, NY, USA

Abstract

People's opinions are one of the defining factors that turn spaces into meaningful places. Online platforms such as Yelp allow users to publish their reviews on businesses. To understand reviewers' opinion formation processes and the emergent patterns of published opinions, we utilize natural language processing (NLP) techniques especially that of aspect-based sentiment analysis methods (a deep learning approach) on a geographically explicit Yelp dataset to extract and categorize reviewers' opinion aspects on places within urban areas. Such data is then used as a basis to inform an agent-based model, where consumers' (i.e., agents') choices are based on their characteristics and preferences. The results show the emergent patterns of reviewers' opinions and the influence of these opinions on others. As such this work demonstrates how using deep learning techniques on geospatial data can help advance our understanding of place and cities more generally.

2012 ACM Subject Classification Computing methodologies → Natural language processing; Computing methodologies → Agent / discrete models; Social and professional topics → Geographic characteristics

Keywords and phrases aspect-category sentiment analysis, consumer choice, agent-based modeling, online restaurant reviews

Digital Object Identifier 10.4230/LIPIcs.GIScience.2023.81

Category Short Paper

Supplementary Material *Software (Source Code)*: `https://github.com/wang-boyu/yelp-abm`
archived at `swh:1:dir:e6398b2a2185fab1b5168bfd588d75261bac2df1`

1 Introduction

People's opinions about places reflect their emotional attachment to locations that hold meanings to them. With the rise of social media platforms such as Google Reviews, TripAdvisor, and Yelp, vast numbers of opinions about local businesses, including restaurants, have been published online. These text reviews provide valuable insights into various aspects of the dining experience, such as the quality of food and service. Studying these reviews through aspect-based sentiment analysis (ABSA) can help identify key aspects that customers care about and understand the rationales behind customers' needs and preferences. The present work aims to address the following research questions (RQ): How to utilize recent advancements in natural language processing (NLP) techniques to: 1) help identify key aspects that customers care about when choosing restaurants, and 2) help inform consumer choice modeling in the context of visitation patterns to restaurants?

Over the last several decades, a body of literature has grown with respect to studying consumers' choice factors when choosing which restaurant to visit. For example, Auty [2] identified several main choice variables (e.g., food type, food quality, location, etc.) in

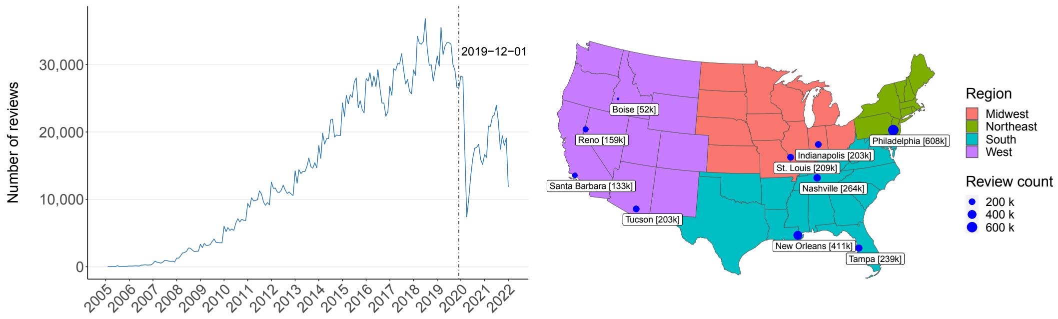
[1] corresponding author

the restaurant selection processes, and how their relative importance varied with respect to consumers' demographic segments such as age and income. Focusing on quick-casual restaurants, Ryu and Han [10] analyzed the relationships between restaurant qualities (i.e., food, service, and physical environment) and consumers' perceived price, and how they in turn affected customer satisfaction and their behavioral intention. Similarly, Bujisic et al. [3] investigated the interactions between restaurant qualities and their effects on consumers' intentions. However, most of the existing studies have used qualitative methods (e.g., surveys, interviews, focus groups) to collect consumer responses. As a result, they are limited by small sample sizes of a few hundred people, and the scope of these studies are either towards specific demographic groups (e.g., students, senior citizens) or specific types of restaurants (e.g., fast-food, Chinese, Korean) [8].

More recently, there has been an emergence of computational social science (CSS) which aims to analyze social phenomena through computational approaches [4]. Accompanying the abundance of digitized text data, one approach in CSS is to develop and utilize new algorithms and toolkits to conduct automated content analysis (e.g., [9]). Sentiment analysis, in particular, aims to automatically estimate or extract subjective sentiments that are expressed through texts, and it can be conducted at different levels. For instance, one may be interested in examining the overall sentiment from an article, or break down an article into sentences or words with their sentiments analyzed separately [7]. Methods that have been utilized to perform such analysis fall into two broad categories: dictionary-based and machine learning.

Dictionaries (or sometimes referred to as lexicons) contain a selected set of words with associated pre-defined sentiment scores. However, one drawback of this category of text analysis is that word order is often neglected. However, it may be crucial for certain types of sentiment analysis, especially at the aspect-level. For instance, an online restaurant review (e.g., from Yelp) may contain several aspects with different sentiments (e.g., good food but poor service) in a single sentence. Ignoring word order may not accurately estimate these aspect-based sentiments. Machine learning methods overcome this issue by encoding and processing texts in a sequential manner, where useful information about word order can be retained. As a result, the performance of machine learning models are often superior to dictionaries, as observed in several studies (e.g., [11]). Deep learning models in particular, have gained popularity in terms of ABSA over recent years [6]. Much work has been carried out to develop new models focusing on improving their predictive power, but they often fall short of advancing theory in explaining and understanding why and how people produce sentiments. It is our aim in this work to demonstrate how these predictive models can be used to investigate salient factors driving peoples' sentiments and link with existing results from qualitative studies, using online restaurant reviews as a case study.

Another approach in CSS is using agent-based modeling (ABM) to simulate complex systems by modeling individual agents and their interactions. In recent years, there have been a trend of integrating machine learning algorithms in and for agent-based models [5]. For example, dimensionality reduction and clustering algorithms have been utilized to analyse model outcomes. Machine learning models have also been used to train on human behavior data and subsequently represent agents during model executions. Agents may collaboratively optimize context-specific goals under the reinforcement learning framework. In order to answer our research question, we combine these two strands of CSS to explore how to utilize deep learning techniques to inform an agent-based model of consumer choices. In what follows, we briefly describe our methodology (Section 2) before presenting the results (Section 3). Finally, Section 4 proved a summary of the paper and identifies areas of further work.



(a) Total number of restaurant reviews by year. **(b)** Number of restaurant reviews in selected cities.

Figure 1 Temporal (a) and spatial (b) distributions of restaurant reviews.

2 Methodology

The Yelp dataset is publicly available at `https://www.yelp.com/dataset` and contains more than 6 million text reviews on over 150,000 businesses in the United States from 2005 to 2021. In this study, we focus on a sampled set of reviews on restaurants prior to the outbreak of COVID-19. Our rationale for this is that COVID-19 has substantially altered consumer behaviours with respect to visiting restaurants. Figure 1a shows the total number of reviews over time while Figure 1b shows the number of reviews in representative cities.

We use a NLP approach to conduct ABSA on these Yelp restaurant reviews. Following the standard text pre-processing procedures in NLP, sampled texts are pre-processed to remove stop words, punctuation, and so on. Next, we draw on theories from computational linguistics and machine learning (e.g., [1]) to extract and categorize salient aspect terms from the text reviews, such as food, service, price, and ambiance, by applying a pre-trained language model from the PyABSA framework [13]. We also assign sentiments (i.e., positive, negative, and neutral) to these categorized aspects. Finally, a linear regression model is used to identify common sentiment patterns and examine how these patterns vary across different aspects.

While we can estimate casual effects of choice factors on star ratings through theoretical and statistical models, the results are only at an aggregate level, and it is difficult to gain insights on how they may differ for different consumer segments. This is mainly due to the lack of individual data which is a consequence of privacy and ethical concerns. As such, we turn to agent-based modeling to simulate artificial, heterogeneous agents (i.e., consumers) and their restaurant visiting patterns. During model initialization, restaurants are created with locations using information from Yelp, along with their average sentiment scores from the pre-trained language model mentioned above. Consumer agents are created at random places with a random attribute (i.e., student, middle-aged, or senior), which subsequently determines their restaurant preferences. For example, student agents are more sensitive to the price factor, whereas senior agents prefer restaurants with higher ambience score, following findings from past studies through surveys and interviews [8]. At each step, consumer agents are informed by the NLP model results and visit the best restaurant based on their preferences. We also implement a null model in which consumer agents make random decisions on which restaurant to visit. An overview of the model logic is shown in Figure 2.

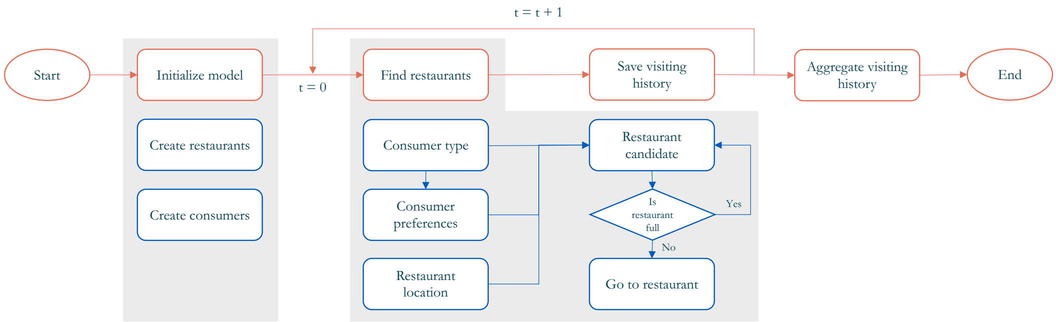


Figure 2 An overview of proposed agent-based model logic.

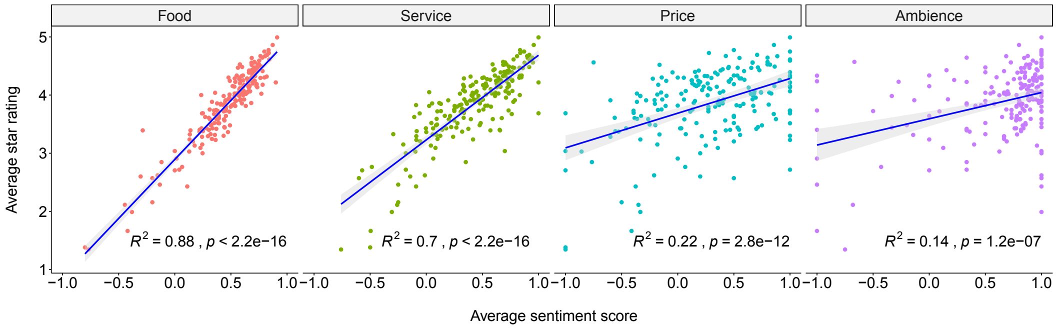
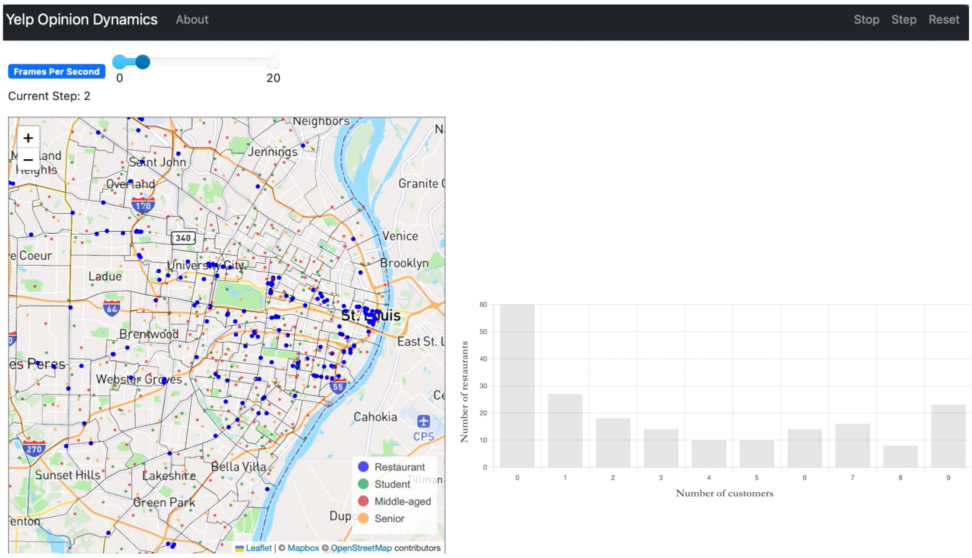


Figure 3 Average star rating vs. average sentiment by aspect category for 200 randomly selected restaurants in the City of St. Louis, MO.

3 Results

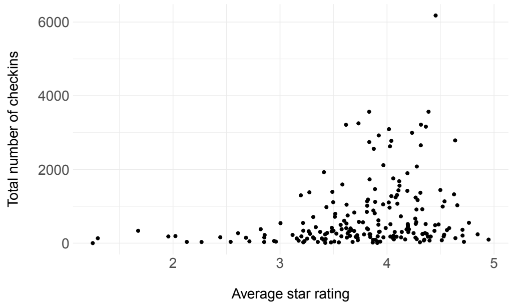
The pre-traind language model [13] transforms each text review and assigns a value to each main aspect category: food, service, price, and ambience, ranging from negative (-1), neutral (0), to positive (1). Although each text review from Yelp also includes a star rating ranging from 1 to 5, this information is never used during the model training. That is, model results are purely based on text data. To help answer RQ 1, we use star ratings as a proxy for consumer satisfaction, and plot average sentiment score of each aspect category against star ratings, for 200 randomly selected restaurants in the City of St. Louis, MO. The results are shown in Figure 3. Unsurprisingly, food appears to be the choice factor that is most strongly correlated with average star ratings while ignoring all other factors. This validates, and at the same time is validated by, previous qualitative studies through surveys and interviews [8]. Notably all results in Figure 3 are statistically significant ($p < 0.001$).

To understand consumers' decision-making processes and address RQ 2, we create a prototype model for the City of St. Louis using the Mesa framework in Python and its GIS extension Mesa-Geo [12]. A screenshot of this prototype model is shown in Figure 4a. Figure 4b shows the average results of 100 simulation runs, and compares it to the null model. Figure 4c shows the actual number of check-ins from Yelp. However a direct comparison between check-ins and visits is difficult to make because not all Yelp users do check-ins on each visit. Our model shows that there are significantly more consumer visits to restaurants with a star rating above 3 than to those with a star rating below 3. For the null model, restaurant visits are evenly distributed regardless of star ratings, which is to be expected. To some extent, this is also reflected in the actual number of check-ins versus actual star ratings.



(a) Screenshot of the graphical user interface of the proposed agent-based model.

(b) Average number of consumers vs. star rating results from the simulations.



(c) Actual number of check-ins vs. star rating from the Yelp data.

Figure 4 The prototype agent-based model (a) with simulated (b) and actual visiting patterns (c).

4 Summary and Areas of Further Work

Online customer reviews can provide valuable insights into various aspects of people's dining experience, such as the quality of food and service. In this paper we have utilized ABSA methods on the Yelp dataset to extract and categorize reviewers' opinions on restaurants in urban areas. Theses estimated opinions form a basis for subsequent statistical analysis and simulation through agent-based modeling. Within the context of this paper we see several areas of further work. A potential area to be further improved regarding sentiment analysis is to experiment with alternative language models of higher predictive performance, and fine-tune such models with more restaurant review data. In terms of the agent-based model, there is always room to extend and refine them. The first relates to incorporating more census data into the model when initializing the consumer agents in order to better stylize and build our synthetic population. It would also be interesting to explore a more finer time granularity that would capture different parts of the day such as mornings, afternoons and evenings as this might also impact visitation to different types of restaurants. Lastly, efforts could be made to calibrate model parameters and validate model results with restaurant check-in data from Yelp or other check-in data sets such as Google. Even with these areas of further work, this paper demonstrates how using deep learning techniques can help advance our understanding of people's choices when it comes to visiting various locations within a city and how such analysis can be incorporated within agent-based models to explore how people interact with places and influence each other.

References

1 Eman Saeed Alamoudi and Norah Saleh Alghamdi. Sentiment classification and aspect-based sentiment analysis on yelp reviews using deep learning and word embeddings. *Journal of Decision Systems*, 30(2-3):259–281, 2021. doi:10.1080/12460125.2020.1864106.

2 Susan Auty. Consumer choice and segmentation in the restaurant industry. *Service Industries Journal*, 12(3):324–339, 1992. doi:10.1080/02642069200000042.

3 Milos Bujisic, Joe Hutchinson, and Haragopal Parsa. The effects of restaurant quality attributes on customer behavioral intentions. *International Journal of Contemporary Hospitality Management*, 26(8):1270–1291, 2014. doi:10.1108/IJCHM-04-2013-0162.

4 Claudio Cioffi-Revilla. *Introduction to computational social science*. Springer, New York, NY, 2017. doi:10.1007/978-3-319-50131-4.

5 Andrew Crooks, Alison Heppenstall, Nick Malleson, and Ed Manley. Agent-based modeling and the city: A gallery of applications. In Wenzhong Shi, Michael F. Goodchild, Michael Batty, Mei-Po Kwan, and Anshu Zhang, editors, *Urban Informatics*, pages 885–910. Springer, Singapore, 2021. doi:10.1007/978-981-15-8983-6_46.

6 Hai Ha Do, Penatiyana WC Prasad, Angelika Maag, and Abeer Alsadoon. Deep learning for aspect-based sentiment analysis: a comparative review. *Expert Systems with Applications*, 118:272–299, 2019. doi:10.1016/j.eswa.2018.10.003.

7 Bing Liu. *Sentiment analysis: mining sentiments, opinions, and emotions*. Cambridge University, Cambridge, UK:, 2015. doi:10.1017/CBO9781139084789.

8 Caroline Opolski Medeiros, Elisabete Salay, et al. A review of food service selection factors important to the consumer. *Food and Public Health*, 3(4):176–190, 2013. doi:10.5923/j.fph.20130304.02.

9 Ross Petchler and Sandra González-Bailon. Automated content analysis of online political communication. In Stephen Coleman and Deen Freelon, editors, *Handbook of digital politics*, pages 433–450. Edward Elgar Publishing, Cheltenham, UK, 2015. doi:10.4337/9781782548768.00037.

10 Kisang Ryu and Heesup Han. Influence of the quality of food, service, and physical environment on customer satisfaction and behavioral intention in quick-casual restaurants: Moderating role of perceived price. *Journal of Hospitality & Tourism Research*, 34(3):310–329, 2010. doi:10.1177/1096348009350624.

11 Wouter Van Atteveldt, Mariken ACG Van der Velden, and Mark Boukes. The validity of sentiment analysis: Comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms. *Communication Methods and Measures*, 15(2):121–140, 2021. doi:10.1080/19312458.2020.1869198.

12 Boyu Wang, Vincent Hess, and Andrew Crooks. Mesa-Geo: A GIS extension for the Mesa agent-based modeling framework in python. In *Proceedings of the 5th ACM SIGSPATIAL International Workshop on GeoSpatial Simulation, (GeoSim '22)*, page 1–10. Association for Computing Machinery, 2022. doi:10.1145/3557989.3566157.

13 Heng Yang and Ke Li. PyABSA: Open framework for aspect-based sentiment analysis. *arXiv preprint arXiv:2208.01368*, 2022. doi:10.48550/arXiv.2208.01368.