國立臺灣大學管理學院資訊管理學研究所

碩士論文

Department of Information Management

College of Management

National Taiwan University

Master Thesis

基於使用者生成內容進行餐廳選址排序

RESPECT: Restaurant Placement Ranking via Leveraging User-Generated Contents

江雨柔

Yu-Jo Chiang

指導教授：魏志平 博士

Advisor: Chih-Ping Wei, Ph.D.

中華民國 112 年 7 月

July 2023

doi:10.6342/NTU202303218

致謝

時光荏苒，碩二的時間走的飛快，首先感謝在我茫然無措進入資管所時，將我納入羽下的魏志平教授。老師真的就像大家的大家長一樣，時不時關心以及帶領我們一起學習或是研究。耐心的教導、不厭其煩地叮嚀，以及24小時隨時解答論文研究的疑惑，真的是我進資管所最大的幸運。最重要是老師總是能從容不迫的解決很多問題，讓常常手足無措不知如何是好的我吃了不少定心丸，還有偶而也會開示我們人生還有很多選項不只是窮忙窮碌。

感謝我的同學們：海因、佑甄、佩琳、宇雋以及璟耀。兩年一起努力的感覺真好，碩一一起為了課業、實習在實驗室拼搏，碩二為了碩論埋頭苦幹，當然還有那些快樂的小確幸時光，讓研究所的路沒有那麼辛苦，現在回想都覺得奢侈又幸福，很高興能認識一群又棒又厲害的小夥伴們。特別感謝博班的學長們以及大學長：仁瀚、晉華與沛原學長，在我寫論文這段期間給予了莫大的幫助，不論是方法的討論或是單純解惑。尤其仁瀚學長真的是盡心盡力，讓我的研究每次卡住的時候總是又能找到下一個出口，謝謝學長花如此多的時間和我一起討論一起分析，讓我的論文得以邏輯通暢且完整。

感謝我的家人們，無私的栽培讓我得以無憂無慮地完成學業。感謝我的小老妹，默默的支持與守護讓我求學時能心無旁騖的把我的這份學業完成。感謝我周遭的新朋友、老朋友們陪我、鼓勵與支持以及一起完成研究所學業，研究所固然辛苦，但也因為有你們而顯得值得。

最後，開始著手進行研究到最後完成這份論文，是我截止目前為止最浩大的工程，要感謝的人絕對遠大於上述，在此統一感謝在這一路上給我信心、見解、指導的同學、老師、朋友、同事與上司，讓我這一路走來真的不孤單。回顧至今，這些磨練與經歷都讓我更上一層樓；這些陪伴都讓我更加成長茁壯。謝謝大家一起見證，未來將會帶著這份祝福與鼓勵繼續努力。

江雨柔 謹識

于臺灣大學資訊管理學研究所

中華民國一百一十二年七月

doi:10.6342/NTU202303218

摘要

在餐飲業日益普及的背景下，如何在百家爭鳴中脫穎而出成為重要的商業議題，策略性地選址規劃變得格外關鍵。除了從供給與需求數據作為決策考量外，在社群媒體發展快速的年代，分析使用者生成的內容 (UGC) 來了解消費者在乎的面相以及對應的滿意度更是重要的資訊管道。透過屬性層級的情感分析（ABSA）技術的不斷改進，我們現在能夠更快且更真實地去了解用戶的反饋與需求。

在預測新店面適合的位置時，我們使用了多種特徵，其中包括描述候選位置以及餐廳的表現和競爭力的相關資訊。然而，傳統的競爭力計算方法卻常常依賴於商業資料、商店密度或評分等單一因素。為了克服這些限制，我們提出了一種新的方法，採用了 ABSA 技術來分析客戶的評論和反饋。此外，我們的競爭者分析採用了社群偵測技術，使得競爭者群組得以更為準確地被識別。這種方法與傳統的基於類別的競爭者分析相比，能夠提供更精確和深入的競爭環境分析。

經過一系列實驗，我們的研究結果顯示我們所提出的方法(稱為 RESPECT)的有效性，並進一步解釋所得出的分析結果。我們所提出的 RESPECT 方法不僅能夠顯著改進餐廳選址決策的準確性和相關性，還展現了利用 ABSA 技術來制定更細緻、更數據驅動策略的潛力。總結，本論文的貢獻在於提出了一種全新且高效的方法，將情感分析和競爭者分析有機地結合，並採用 ABSA 技術來實現對餐廳選址的精細化評估。我們的研究將為餐廳業務和選址規劃提供寶貴的洞察，幫助餐廳業主和相關利益相關者做出更明智的決策，以在競爭激烈的行業中脫穎而出。

關鍵字：新店家選址、餐廳選址排序、社群分析、使用者生成內容、屬性層級情感分析、深度學習

doi:10.6342/NTU202303218

ABSTRACT

In the context of the increasingly popular catering industry, how to stand out in the intense competition has become essential; hence, selecting optimal locations for opening brick-and-mortar restaurant branches becomes even more indispensable. Besides relying on supply and demand data as predictive factors, with the proliferation and increasing popularity of social media platforms, user-generated contents (UGC) has increasingly become an important information source to understand consumers' preferences and satisfaction. With the use of aspect-based sentiment analysis (ABSA) methods, we now have a faster and more authentic means of comprehending user feedbacks and demands.

In predicting suitable locations for new restaurant establishment, we employ multiple features, descriptions of candidate locations, restaurant performance, and competitiveness-related information. However, conventional methods rely on limited data sources, such as commercial data, shop density, or ratings. To overcome these limitations, we propose a novel approach that analyzes customer reviews and feedbacks. We also utilize review with community detection to offer more precise and in-depth analysis of the competitor landscape.

Our research results demonstrate the effectiveness of our proposed method (referred to as RESPECT). RESPECT not only significantly enhances the accuracy restaurant placement decisions but also showcases the potential of using ABSA techniques to formulate more refined and data-driven strategies. In conclusion, the study offers valuable insights for restaurant businesses and site planning, aiding restaurant owners and relevant stakeholders in making more informed decisions to excel in the fiercely competitive industry.

Keywords: New store location selection, Restaurant placement ranking, Community detection, User generated content, Aspect-based sentiment analysis, Deep learning

doi:10.6342/NTU202303218

Table of Contents

doi:10.6342/NTU202303218

List of Tables

List of Figures

doi:10.6342/NTU202303218

Chapter 1 Introduction

1.1 Background

Location of a store is a crucial factor in determining its success and profitability when opening a new branch. As the age-old maxim goes: “No matter how good it is offering, merchandising, or customer service, every retail company still has to contend with three critical elements of success: location, location, and location”(Taneja, 1999). Surveys conducted by the Small Business Administration (SBA) and other organizations indicate that poor location is one of the primary causes of business failure in America[1]. Fox et al. (2007) and Li and Liu (2012) all state that even though there are many factors influencing the success or failure of a store, one factor still manages to stand out, which is location. A survey conducted by Foursquare[2] states that when the store location is not accurately determined, 46% of the businesses gain higher customer acquisition costs, 40% lose market shares, and 42% has lower level of customer experience.

It is acknowledged that the level of success of a business is heavily dependent on operating the business from the right location. The key determinant is demand, hence, defining an attractive site location that will generate high demand and would minimize costs is the ultimate goal. Shop owners or managers will want to choose a location where people have the inclination to shop, the time to spend in your store, and the propensity to buy your product. Features can include the state of the environment, the number and location of competitors as well as the degree of their success, the cost of renting the premises, and a wide range of indicators related to future attendance of the establishment

[1] Dirkes, D. (2014) Business Failure Stats & Reasons Summary. Linkedin (Sep 21, 2014) Available at https://www.linkedin.com/pulse/20140921225654-8175758-business-failure-stats-reasons-summary (Retrieved on July 23, 2023).

[2] Forrester (2022) Location Intelligence Builds Innovation, Optimization, And Better Customer Experiences. (July, 2022) Available at https://location.foursquare.com/wp-content/uploads/sites/2/2022/08/Forrester_Opportunity_Snapshot_-_Report.pdf (Retrieved on July 23, 2023).

doi:10.6342/NTU202303218

(Hernandez & Bennison, 2000). In conclusion, new store location selection is a complex and crucial process that requires comprehensive consideration of various factors. Through scientific analysis and strategic planning, we will be able to identify the ideal location best suited for store development.

Following are some real-world examples demonstrating how business nowadays make decision on deciding optimal new locations. Since our main focus is new store establishment for restaurants, we narrow our business down to catering industry. By the end of 2022, restaurants with the most branches are as follows: McDonald's has around 40,275 branches worldwide[3], Starbucks with 35,711 branches[4], and KFC with over 27,700 branches[5]. With so many stores to choose from, this is not an easy task to do. Looking at Figure 1-1, they have managed to maintain their position within the top three in terms of brand value.

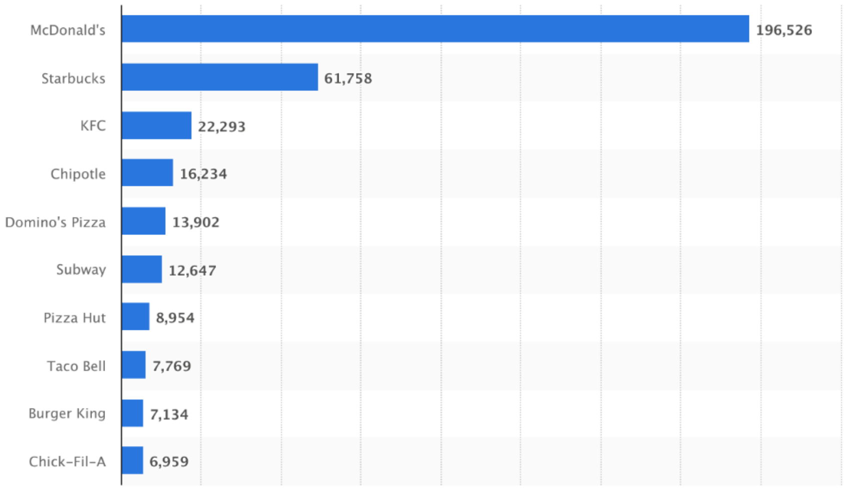


Figure 1-1: 10 Most Valuable Quick Service Restaurant Brands Worldwide in 2022[6]

[3] McCain A. (2023) 22 McDonald's Statistics [2023]: Restaurants Counts, Facts, and Trends. Zappia (March 21, 2023) Available at https://www.zippia.com/advice/mcdonalds-statistics/ (Retrieved on July 23, 2023).

[4] Statista Research Department (2023) Number of international and U.S.-based Starbucks stores from 2005 to 2022. Statista (Nov 28, 2022) Available at https://www.statista.com/statistics/218366/number-of-international-and-us-starbucks-stores/ (Retrieved on July 23, 2023).

[5] Statista Research Department (2023) Number of Kentucky Fried Chicken restaurants worldwide 2010-2022 Statista (Mar 31, 2023) Available at https://www.statista.com/statistics/256793/kfc-restaurants-worldwide-by-geographic-region/ (Retrieved on July 23, 2023).

[6] Statista Research Department (2023) Brand value of the 10 most valuable quick service restaurant brands worldwide in 2022 Statista (May 9, 2023) Available at https://www.statista.com/statistics/273057/value-of-the-most-valuable-fast-food-brands-worldwide/ (Retrieved on July 23, 2023).

doi:10.6342/NTU202303218

McDonald’s has an extensive and detailed decision system for how to locate stores, which includes demographic databases and population statistics-based marketing research to target consumers and analyze competition within a radius of five to seven miles and find out if the location is suitable for franchises. KFC also sets up an estimation method to score the locations or business districts by its population, consumption power, revenue of the business district and the convenience of transportation. After evaluating through all these metrics, KFC chooses the highest-scored places and distributes branches in buildings, supermarkets, bookstores, trade centers, etc. Atlas provides Starbucks with data on consumer demographics, population density, income levels, auto traffic patterns, public transport stops, and the types of stores/businesses at the location under evaluation. Thus, Starbucks is able to project foot traffic and average customer spending at a given location, therefore helping Starbucks to determine the economic viability of opening a store in that spot. In general, past literature and businesses nowadays confirm a statistically significant relationship between site location and business performance, and continuing efforts have been made to determine emerging factors affecting this relation.

The problem of new store location selection has been extensively studied. When opening a new store or branch, many factors and decisions should be considered. By investigating through the existing literature, we can separate these methods into two approaches: **survey-based** and **data-analytics-based**. Survey-based methods are those that make use of surveys and expert experiences, which are too time-consuming and need to be more scalable in this fast-changing world. In contrast, data-analytics-based methods collect data from all sources, retrieving data with both volume and velocity. The features employed by these methods can further be divided into three groups according to their data sources, namely **geo-economics-based**, **location-based**, and **social-media-based** features. Geo-economics-based features take advantage of geographic, demographic, and

doi:10.6342/NTU202303218

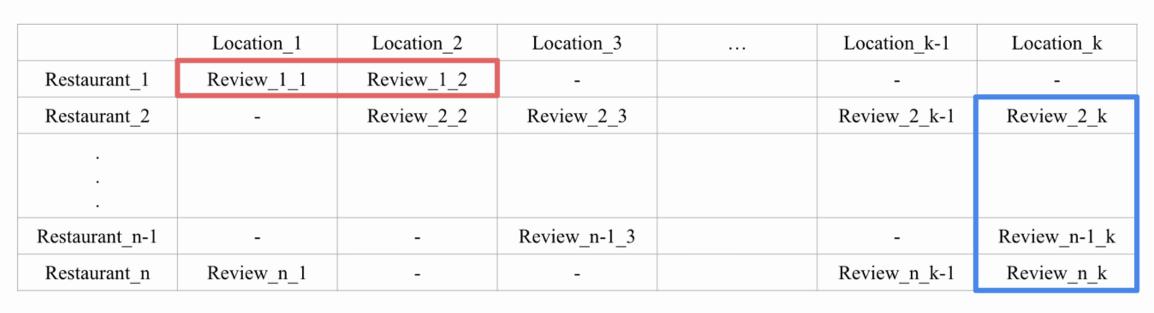
transportation perspectives to describe places or restaurants from consumers' perspectives rather than just features relevant to store or restaurant itself. A clearer picture of consumer behavior at a specific location, allows us have a better understanding of consumers and segmentation, leading to a more accurate prediction. Location-based features include data from mobile apps or social media platforms, such as Foursquare or maps' point of interests (POI) check-ins depicting the factors with customer flow showing the popularity of one location with spatial and temporal information. Most of the works tackle the problem with this approach (Fox et al., 2007; Li & Liu, 2012). Social-media-based features require data from social media applications or review websites. By leveraging the user-generated contents, we can now take users' opinions into consideration. We can also gain insight into social validation such as likes and comments. Such validations not only signal customers but also affect the behavior of other customers. The social-media-based features typically focus on capturing information related to a user's location, activities, interactions, and preferences, scaling the restaurant from the consumer perspective, so that we can have a finer view of the performance of restaurants.

1.2 Research Motivation and Objectives

The existing studies that consider reviews as an input typically view review counts as a way to measure popularity or success, instead of using the aspects or sentiments inside them. Reviews are important sources to gain consumer preferences, and once we identify what consumers like about our restaurant, we can than find places that show similar preference to enhance the successful rate of opening a new branch. The reviews used to represent the restaurant are collected from all of its branches. Assume that restaurant_1 now have stores at location_1 and location 2 and we are deciding whether restaurant_1 should open its new store at location_k. In this case, we use the reviews about restaurant_1

doi:10.6342/NTU202303218

at location_1 and location_2 (i.e., the red area shown in Figure 1-2) to represent restaurant_1. On the other hand, we aggregate the reviews of all other restaurants that have opened at location_k to model location_k (i.e., the blue area shown in Figure 1-2). Using the reviews from these two areas, we can estimate the appropriateness of opening a restaurant_1 store at this specific location (i.e., location_k).



	Location_1	Location_2	Location_3	…	Location_k-1	Location_k
Restaurant_1	Review_1_1	Review_1_2	-		-	-
Restaurant_2	-	Review_2_2	Review_2_3		Review_2_k-1	Review_2_k
. . .						
Restaurant_n-1	-	-	Review_n-1_3		-	Review_n-1_k
Restaurant_n	Review_n_1	-	-		Review_n_k-1	Review_n_k

Figure 1-2: Reviews for Branches of Different Restaurants

Chapter 2 A study by Wang et al. (2016) is the first to leverage aspects and sentiment of reviews to gain consumer preference. The aspects extracted from Latent Dirichlet Allocation **(**LDA) are at the document level. Thus, the topics extracted do not correspond to aspects. Furthermore, the LDA topics are presented by the words in documents, resulting in too fine grained or diverse aspects. Accordingly, we propose in this study a **REStaurant PlacEment ranking model via leveraging user-generated ConTents (RESPECT)**, which includes LDA extracted aspects, pre-defined aspects and ASBA-extracted aspects to capture more fine-grained aspects from consumer reviews. Also, the sentiments of the aspects discussed in reviews are highly valued in social media and marketing fields. Other than adding up the sentiment score and viewing the performance as a whole, we can combine the sentiment score with aspect importance to create a new way to evaluate performance of a restaurant at the aspect level. In addition to sentiment information, reviews can also disclose competitor relations, such as the degree of competitors clustering. These novel features mentioned above can offer a new perspective

doi:10.6342/NTU202303218

to tackle the new store location selection problem.

As for defining new store location selection, most of the work focuses on predicting the ratings or revenue of one restaurant to represent the success of opening at a specific location. By providing rankings of location candidates, we not only simulate the real-world selection process, but also reveal additional information to decision makers. Therefore, this study develops a ranking model based on geo-economics-based features, location-based features, and social-media-based features to find the best restaurant locations. We examine the following research questions in this research:

- Can sentiment analysis help to generate more useful features?
- Does the aspects extracted from different ways affect ranking effectiveness?
- Apart from sentiment analysis, can the degree of competitors feature extracted from reviews contribute to more effective restaurant placement?

doi:10.6342/NTU202303218

Chapter 2 Literature Review

2.1 Existing New Store Location Selection Methods

When dealing with the new store selection problem, most of the adopted features mainly characterize the characteristics of location and store. The data source of the feature groups can be divided into three main groups as mentioned earlier, including geo-economics-based features, location-based features, and social media-based features. In recent years, the introduction of Geographic Information Systems (GIS), location-based applications like map search, navigation systems, and social media applications into the world has led to an increase of relevant studies employing location-based features, for example, point of interest (POI) check-in data and social-media-based features like user-generated contents (UGC). As can be seen from previous studies, different feature sources have been adopted simultaneously, for example, there has been a gradual shift from the use of only geo-economics-based features to the use of a mixture of the above features. Though sources change over years, the same concepts are still being captured. For the feature group for location, the overall concept is to describe whether the chosen location is suitable to withhold a new store opening, including demand, supply, traffic accessibility and the consumption power of the location. Regarding the feature group for stores, more focus is on capturing the performance and the competitiveness of the store itself, for example, willingness to pay of the customers, ratings, profit, etc.

2.1.1 Feature Group for Location

Supply of an area is usually depicted by shop density and demand can be captured by human flow or population in that area. Most works of the geo-economics-based

doi:10.6342/NTU202303218

features focus on census data such as population, birth rate, mortality rate, crime rate, immigration rate, replacement rate, etc. These features allow us to get a picture of that neighborhood's supply and demand in quantitative measures. Location-based features such as check-in data and web search data provide insights of human flow. With the knowledge of these foot traffic, it is easier to choose the hot spot of the neighborhood. Karamshuk et al. (2013) used the dataset collected from Foursquare in New York. They employ and develop geographic and mobility features that serve as demand for popularity prediction on three different coffee shops and restaurant chains (i.e., Starbucks, Dunkin' Donuts, and McDonald's) in New York City. Rahman and Nayeem (2017) also take advantage of Foursquare New York City to estimate demand and the capability of gathering audiences and used them to predict a location for live campaigns. Xu et al. (2016) detect the gaps between demands and supply by extracting timestamps and locations through queries from Baidu search engine. These extracted keywords are then transformed into a demand point to specify the temporal and spatial features of demands. Zhdanov et al. (2022) collect reviews from TripAdvisor to gain insight into restaurant distribution in Russia.

Most studies indicate that easy access to public transportation is an important factor that influences consumers to travel to the place. When it is more difficult to reach a destination, consumers are less likely to go there (Ge et al., 2019; Sokol & Jordanov, 2020; Yu et al., 2016). In most cases, traffic convenience is a necessary but not sufficient factor to the success of a shop. Some studies take distance to downtown and traffic accessibility as indicators to measure the transportation convenience (Li et al., 2021; Morrison & Abrahamse, 1996). Other studies use maps and POI check-in data to capture the availability of public transportations (Guo et al., 2018; Rahman & Nayeem, 2017; Sokol & Jordanov, 2020; Zeng & Tang, 2019).

doi:10.6342/NTU202303218

Morrison and Abrahamse (1996) take account of future consumption base of a neighborhood and specific demographic factors for new store location selection for a large supermarket catering to one-stop shoppers. Geo-economics-based features take household income, house price to represent the consumption power of a certain location, while location-based and social-media-based features offer valuable information about a user's geographical context interests and preferences, they generally do not directly reveal information about a user's financial status, income, or wealth.

2.1.2 Feauture Group for Stores

Competitiveness and complementary are two main features that most prior studies focus on. Geo-economic-based features take shop commercial profile, expertise insights to measure a brand's competitiveness (Morrison & Abrahamse, 1996), while types created by users can be found in the source of location-based and social-media-based. These types can also be an input to calculate shop diversity. A handful of studies apply neighborhood entropy measure proposed by Karamshuk et al. (2013) by using location-based features (Sokol & Jordanov, 2020; Yu et al., 2016; Zeng & Tang, 2019). Additionally, social-media-based features, to be precise, UGCs, offer compatibility to an location and also account for measure competitiveness. Wang et al. (2016) utilize the comments by using LDA model to evaluate restaurants at an aspect level and calculate the compatibility with the location.

The competition or cooperation between shops in the neighborhood may greatly affect the success rate of the new entering shop. Jensen Quality index (Jensen, 2006) is widely used in calculating interactions between venues. Geo-economics-based features use commercial profiles to calculate the index, while location-based features and social media use shop types to perform this calculation. Co-occurrence of venues in an area is

doi:10.6342/NTU202303218

also viewed as complementary, the higher frequency of different stores locating at the same location might imply a shared advantage cooperation. Yu et al. (2016) model the complimentary and competition between shops, and Ahedo et al. (2021) model the balanced tenancy in an area, using location-based features to compute. Social-media-based features are also used. For example, Wang et al. (2016) use reviews to extract aspects as previously stated to calculate complementary.

2.1.3 Summary of Existing New Store Location Selection Methods

Table 2-1 and Table 2-2 provide a summary of location feature group and store feature group used by prior studies. From these tables, we can observe that the source type of features has become hybrid instead of just using a single source type, meaning that more diverse data are taken into thought patching together a holistic view of the store or location. We can also observe that both location-based and social-media-based features contain more information in a single data source; hence, less data sources are needed to construct the feature groups. Another worth noting is that, for location feature groups, supply and demand are the most captured features and both competitiveness and complementary are adopted for characterizing store details.

doi:10.6342/NTU202303218

Table 2-1: Summary of Feature Groups for Location

Features			Studies									
Feature Group	Source	Source Type	Morrison and Abrahamse (1996)	Karamshuk et al. (2013)	Xu et al. (2016)	Wang et al. (2016)	Rahman and Nayeem (2017)	Guo et al. (2018)	Zeng and Tang (2019)	Sokol and Jordanov (2020)	Li et al. (2021)	Zhdanov et al. (2022)
Supply	Commercial profile	Geo-economics	V	V	V							
	POI check-in	Location		V			V	V	V	V	V	V
	UGC	Social Media			V	V					V	V
Demand	Census data	Geo-economics	V		V	V						
	POI check-in	Location		V			V	V	V	V	V	V
	web search				V							
	UGC	Social Media			V						V	V
Transpor-tation	Traffic accessibility	Geo-economics	V		V							
	Distance to downtown				V							
	POI check-in	Location					V	V	V	V		
Con-sumption Power	Household income	Geo-economics	V		V		V					
	House price							V				

Table 2-2: Summary of Feature Groups for Store

Features			Studies									
Feature Group	Source	Source Type	Morrison and Abrahamse (1996)	Karamshuk et al. (2013)	Xu et al. (2016)	Wang et al. (2016)	Rahman and Nayeem (2017)	Guo et al. (2018)	Zeng and Tang (2019)	Sokol and Jordanov (2020)	Li et al. (2021)	Zhdanov et al. (2022)
Competi-tiveness	Commercial profile	Geo-economics	V		V			V				
	Expertise insights		V									
	POI check-in	Location			V	V			V	V	V	V
	UGC	Social Media				V					V	V
Comple-mentary	Commercial Profile	Geo-economics		V	V	V		V		V	V	
	POI check-in	Location		V				V		V	V	V
	UGC	Social Media				V						

2.2 Prediction Tasks

Prior studies define the new store location selection problem into two tasks, namely, performance prediction and ranking. Performance prediction task focuses on predicting one restaurant or store’s performance at a time, while ranking takes all candidate locations into consideration and return a ranked list for a focal restaurant. Whether it is a performance prediction or ranking task, the success of a store is mostly valued by popularity. Thus, check-in number is an important source that serves as the dependent variable. Most of prior works generally view the new store location selection problem as a performance prediction, and the performance ranking task is less used in prior studies. Table 2-3 summarizes the existing studies on new store site selection, which differ in how the performance of new stores is defined.

doi:10.6342/NTU202303218

Table 2-3: Summary of Prior Studies on New Store Location Selection

Problem	Label (y)	Studies
Prediction	Revenue	Morrison and Abrahamse (1996); Li et al. (2021); Ismail and Hooy (2023)
	Check-in count	Karamshuk et al. (2013); Wang et al. (2016); Zeng and Tang (2019); Sokol and Jordanov (2020); Alharthi et al. (2020)
	Ratings	Guo et al. (2018)
	Post popularity	Li et al. (2021)
Ranking	Check-in count	Xu et al. (2016); Rahman and Nayeem (2017)
	Ratings + review count + review sentiment	Zhdanov et al. (2022)
	Revenue	Damavandi et al. (2019)

2.3 Summary of Literature Review

To sum up, most prior works apply geo-economics-based features and location-based features, which are restaurant-independent. They only reflect the general area properties of the locations without taking the restaurant properties into consideration. Social-media-based features generate attributes unique to the restaurant itself, depicting restaurants in a more personal view, but most prior studies stop at the phase of rating and review count, instead of using the content itself. Whether it is formulated as a performance prediction or ranking task, check-in count or review count are the most commonly used features to show the successfulness of one restaurant. Hence, we follow this method to design our dependent variable.

Current research gaps are as follow:

- Few works focus on the content of review to extract competitiveness features, ignoring the details of certain aspects when evaluating a restaurant.
- Using LDA model to do aspect extraction results in lack of interpretability, the generated topics may not directly correspond to meaningful or intuitive aspects

doi:10.6342/NTU202303218

that users are interested in. And since LDA assumes that each word in a document is associated with a single aspect, this may lead to difficulty in handling overlapping aspects causing the model to assign words to separate aspects even when they are closely related.

- Sentiment analysis is applied in a general level by summing up the score to represent the restaurant, blurring the strength and weakness of a restaurant in a certain aspect.

The novelty of our proposed RESPECT method is:

- Instead of using LDA to do aspect extraction, we also develop two techniques to extract more precise aspects and evaluate a restaurant's performance in a more comprehensive approach.
- Aspect-based sentiment analysis (ABSA) is applied and calculated as indicators of measuring the compatibility of the restaurant and the location.
- Besides using restaurant types to represent the competitors, we deploy a community detection technique to identify competitor clusters, creating a novel feature for the task.

Table 2-4 and

Table 2-5 showcase the features employed in our proposed method and a comparison with prior studies. Most of the features utilized in our proposed method are social-media-based features, with our primary focus centered on the assessment of demand and supply for location-based feature groups, aligning closely with recent research efforts in this domain.

doi:10.6342/NTU202303218

Table 2-4: Our Proposed Method Compared with Prior Works (Feature Group for Location)

Features			**Studies**										
Feature Group	Source	Source Type	MA (96)	KNS (13)	XMW (16)	WCP (16)	RN (17)	GLZ (18)	ZT (19)	SJ (20)	LCZ (21)	ZGD (22)	**Our method (2023)**
Supply	Commercial profile	Geo-economics	V	V	V								**V**
	POI check-in	Location		V			V	V	V	V	V	V	
	UGC	Social Media			V	V					V	V	
Demand	Census data	Geo-economics	V		V	V							
	POI check-in	Location		V			V	V	V	V	V	V	**V**
	web search				V								
	UGC	Social Media			V						V	V	
Transportation	Traffic accessibility	Geo-economics	V		V								
	Distance to downtown				V								
	POI check-in	Location					V	V	V	V			
Consumption Power	Household income	Geo-economics	V		V		V						
	House price							V					

MA (96): Morrison and Abrahamse (1996)
KNS (13): Karamshuk et al. (2013)
XMW (16): Xu et al. (2016)
WCP (16): Wang et al. (2016)
RN (17): Rahman and Nayeem (2017)
GLZ (18): Guo et al. (2018)
ZT (19): Zeng and Tang (2019)
SJ (20): Sokol and Jordanov (2020)
LCZ (21): Li et al. (2021)
ZGD (22): Zhdanov et al. (2022)

doi:10.6342/NTU202303218

Table 2-5: Our Proposed Method Compared with Prior Works (Feature Group for Store)

Features			**Studies**										
Feature Group	Source	Source Type	MA (96)	KNS (13)	XMW (16)	WCP (16)	RN (17)	GLZ (18)	ZT (19)	SJ (20)	LCZ (21)	ZGD (22)	**Our method (2023)**
Competitiveness	Commercial profile	Geo-economics	V		V			V					
	Expertise insights		V										
	POI check-in	Location			V	V			V	V	V	V	
	UGC	Social Media				V					V	V	**V**
Complementary	Commer-cial Profile	Geo-economics		V	V	V		V		V	V		
	POI check-in	Location		V				V		V	V	V	
	UGC	Social Media				V							

MA (96): Morrison and Abrahamse (1996)
KNS (13): Karamshuk et al. (2013)
XMW (16): Xu et al. (2016)
WCP (16): Wang et al. (2016)
RN (17): Rahman and Nayeem (2017)
GLZ (18): Guo et al. (2018)
ZT (19): Zeng and Tang (2019)
SJ (20): Sokol and Jordanov (2020)
LCZ (21): Li et al. (2021)
ZGD (22): Zhdanov et al. (2022)

doi:10.6342/NTU202303218

Chapter 3 Our Proposed Method: RESPECT

3.1 Problem Formulation

We define the new store location selection task as follows in our study: Given a restaurant with *n* new stores to establish with *m* (where *m* >> *n*) candidate locations to consider, the goal of the task is to have the predicted ranking be as close to the true rank list as possible, which means that the location with actual branch opened in that area should be at the top of the list while those without the branch should be at the bottom. The locations with actual branches opened are ranked by their review count in the true list. We construct various features and integrate them into our ranking model, aiming to predict an optimal ranking list of locations. We exploit different ranking methods including pointwise and listwise to get the best ranking result. Pointwise prediction is considered as a binary classification, while listwise ranking generates a list of ranked locations.

3.2 Overview of Our Proposed Method: RESPECT

As described, we define the new store selection problem as a ranking task and propose our method as a restaurant placement ranking model via leveraging user-generated contents. Data collected from the Yelp dataset are then transformed into three types of features: geo-economics-based features, location-based features and social-media-based features. Geo-economics-based features are those utilizing business profiles to create features measuring the supply and diversity of restaurants at a location, location-based features make use of check-in data to capture the migration of consumers and social-media-based features are processed through aspect-based sentiment analysis

doi:10.6342/NTU202303218

(ABSA) and community detection to extract hidden information from user generated contents and to examine restaurants in a finer measure. Lastly, these features are trained by a ranking model to predict the rank of a list of candidate locations to open a new branch for an existing restaurant brand. The model structure we propose is presented in Figure 3-1.

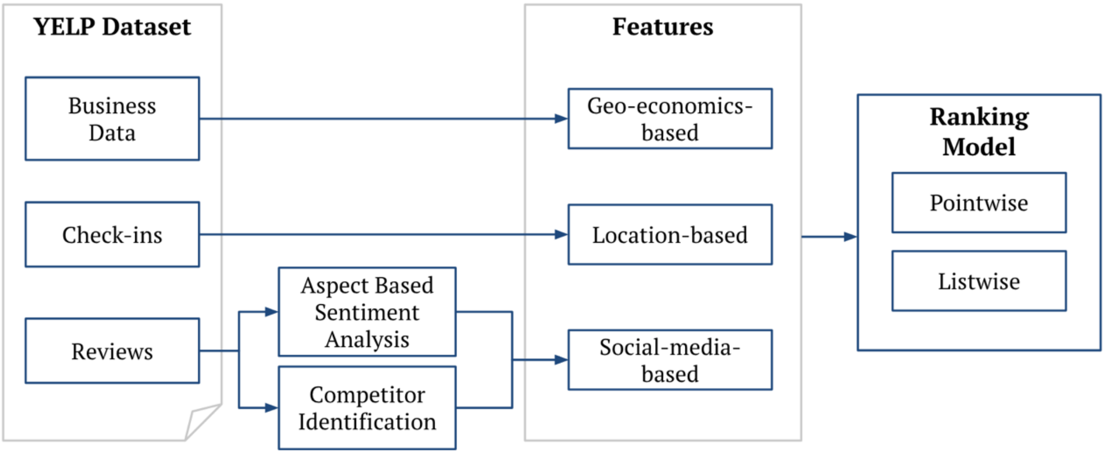


Figure 3-1: Structure of Our Proposed RESPECT Method

3.3 Geo-economics and Location Based Features

We exploit geo-economics-based features and location-based features to capture location and store feature groups inspired by prior works. In our study, we extract these features from the Yelp dataset which includes commercial profiles, check-in data and user-generated contents including ratings, comments and photos.

Density which captures how concentrate one location is with restaurants, where the total number of restaurants located in l is presented as $N(l)$:

$$d_l = N(l) \tag{3.1}$$

The density of one location not only shows the supply but also tells the opportunity or threat in that place in a general way. The higher density in one place shows that more restaurants locate in that area, hence a higher chance to lure people in, but on the downside

doi:10.6342/NTU202303218

meaning there would be an intense competition.

Area popularity shows the visitor amount at one location, where $C(l)$ represents the total check-in count located at location l:

$$c_l = C(l) \tag{3.2}$$

The area's popularity serves as a significant indicator of the demand, playing a critical role in determining the prospects of a restaurant's success. A higher level of popularity not only signifies a greater number of potential customers but also indicates a thriving and bustling environment that can attract a diverse range of patrons. This agglomeration effect, characterized by the concentration of businesses and customers at a specific location, creates a vibrant and competitive atmosphere that fosters synergies and mutual benefits for establishments in the vicinity.

Neighborhood entropy is used to measure the diversity of restaurants in one area, calculated using the entropy measure. Let the entropy be calculated as follows: with the category set Γ, $N_\gamma(l)$ denotes the number of restaurants belonging to the same category γ as restaurant v and located in l. Let the total number of restaurants located in l be $N(l)$. The neighborhood entropy of restaurants at location l is calculated as:

$$ne_{vl} = -\sum_{\gamma \in \Gamma} \frac{N_\gamma(l)}{N(l)} * log \frac{N_\gamma(l)}{N(l)} \tag{3.3}$$

Neighborhood entropy shows the variety of restaurant options at a location, the higher value the more heterogeneous, vice versa. If the location has a low value of neighborhood entropy, it may point out that the location has a preference towards a certain type of restaurants, sending a signal to restaurants with the same type this might be a good place to start or to open a new branch.

doi:10.6342/NTU202303218

3.4 ABSA for Social-Media-Based Features

User-generated contents occupy a large part of social-media-based features. It contains an abundant amount of reviews for just about anything, these reviews serve as an important source of measuring the performance or showing preference towards the commented objects, including store or restaurants. Wang et al. (2016) use the LDA model to extract aspects that serve as criteria to evaluate a restaurant in multiple facets. Their study opens a way for us to measure stores or restaurants in a fine-grained metrics. Following their work, we use different tactics to extract aspects and recreate features to represent the competitiveness and the performance of the targeted restaurant. With the help of aspect-based sentiment analysis (ABSA) methods, we can now analyze restaurants at the aspect level rather than just talking about the overall performance of a restaurant, finding detailed strengths and weaknesses of different restaurants.

Our proposed RESPECT method is based on various ABSA methods to create novel features. The ABSA methods employed are as follows: LDA extracted aspects, pre-defined aspect-guided aspect extraction, and pretrained model for ABSA. Novel features are generated to represent performance of the restaurant and its compatibility with the targeted locations (i.e., the competitiveness of a restaurant at a specific location). In the following subsection, we will introduce the mechanisms used to extract aspects and how the features are being captured.

3.4.1 Latent Dirichlet Allocation (LDA) Extracted Aspects

Latent Dirichlet Allocation (LDA) is a probabilistic generative model commonly used for topic modeling. It assumes that documents are represented as a mixture of topics, and each topic is characterized by a distribution of words. Given a set of restaurants V, restaurant v where $v \in V$, the reviews on restaurant v is viewed as a single document

doi:10.6342/NTU202303218

d. After training the LDA model with all the documents, we can get the probability of a document belonging to a topic k, written as $t_{vk} = p(k|v)$. The goal of LDA is to infer the underlying topic structure of a collection of documents. It works by iteratively assigning words in the documents to different topics and adjusting the topic-word distributions to maximize the likelihood of the observed data, assuming the words in a document are generated based on a probabilistic process involving the topics and their associated word distributions. Before training the LDA model, reviews will go through a series of text preprocessing. First, we tokenize the sentences by spaces, then we do POS tagging and lemmatize all tokens. After that we remove English stop words and numbers, leaving 'nouns' only. Next, tokens occurring more than 50% of documents and less than 50 documents are ignored. Lastly, top 1000 most frequent tokens are kept for LDA training. Once the LDA model is trained, you can access the inferred topic distributions for each document then the aspects are then retrieved using the most important word of the topic and every word represents one topic only. On the basis of the dataset that we collect (to be described in Chapter 4), the retrieved 20 aspects are as follows: "burger, table, beer, taco, order, cream, steak, drink, pizza, sandwich, server, breakfast, wing, sauce, room, chicken, dish, rice, roll, salad".

3.4.2 Pre-defined Aspect-guided Aspect Extraction

After observing the important words generated by the LDA model for 20 topics, we can see that those "topics" are not exactly what we call "aspects". Hence, we carefully choose 30 words as our aspects for analyzing the sentiment of the reviews. The pre-defined aspects are chosen from the Michelin guide and some general assessment indicators in a heuristic matter, which are not only used by inspectors around the world to give out unbiased evaluations to restaurants, but also critics and reviewers around the

doi:10.6342/NTU202303218

world. A total of 30 aspects are chosen, namely 1) food quality, 2) service, 3) environment, 4) value for money, 5) cleanliness, 6) ambiance, 7) menu variety, 8) portion size, 9) presentation, 10) staff Friendliness, 11) speed of service, 12) accuracy of orders, 13) dietary options, 14) drink quality, 15) reservation process, 16) noise level, 17) location, 18) overall experience, 19) authenticity, 20) innovation, 21) beverage selection, 22) wine list, 23) cocktail quality, 24) dessert quality, 25) seating comfort, 26) outdoor seating, 27) special occasion suitability, 28) child-friendliness, 29) parking availability, and 30) online ordering experience.

3.4.3 Pre-trained Model for ABSA

The ABSA task includes several subtasks: aspect term extraction (ATE) and aspect term sentiment classification (ATSC). The goal of ATE is to determine the key topics or aspects being discussed in reviews, enabling a more fine-grained analysis of sentiment or opinion towards those aspects. InstructABSA (Scaria et al., 2023) is a pretrained model we use for extracting aspects in our experiment. The model was trained using the SemEval (Semantic Evaluation) dataset (SemEval 2014 Laptops and Restaurants) and has surpassed previous state-of-the-art methods on downstream ABSA subtasks by a significant margin. The model we use in this study is the ATE model, which returns aspects discussed in a sentence if there are any in the sentence, if not it returns “noaspectterm”. After extracting a total of 209,584 extracted terms from the reviews in our dataset, we perform some filtering and grouping steps to make the aspects more general. The steps are as follows:

I. Lemmatize and remove words that are empty or has ‘restaurant’, ‘naan’ and ‘noaspectterm’ in it.

II. Since most words have document frequency (df) less than 10 with standard

doi:10.6342/NTU202303218

deviation of 100, we filter the aspects by df < 100

III. We then classify the words into six major aspects: ‘food’, ‘drink’, ‘menu variety’, ‘price’, ‘ambience’, and ‘service’. When words cannot be classified into the six aspects, they stand out as new aspects of their own. After processing all words extracted by ABSA, a total of **180 aspects** are captured.

Table 3-1 depicts the number of aspects at each stage of the grouping process, along with the average document frequency, illustrating the diversity of the extracted aspects. Furthermore, the results of decreasing average document frequency indicate that after undergoing the grouping process, these aspects become more generalized.

Table 3-1: Summary of Extracted Aspects

	Original	Lemmatized	Filtered	Grouped
Number of aspect terms	209,584	148,610	944	180
Average document frequency	7.13	9.72	1154.15	4173.29

Some examples of aspects extracted by using the pretrained model for ABSA are illustrated in Table 3-2:

Table 3-2: Example of Extracted Aspects from InstructABSA

	Original Text	**Extracted Aspects**
Example 1	Quite simply, THE go-to restaurant in Clearwater Beach. Try the **grouper sandwich** (Buffalo style) and enjoy their **drink specials**!	[‘grouper sandwich’, ‘drink specials’]
Example 2	We just had the most amazing **meal**! The **environment**, **food**, **staff** were exceptional! We were a large group on a Friday night and it couldn't have been better!	[‘meal’, ‘environment’, ‘food’, ‘staff’]
Example 3	The **Sicilian slice**, fresh out of the oven, was one of the most perfect slices of **pizza** I've had in America. The man who runs this place is a Spanish-speaking Italian grandfather. I want the calzones for birthdays, Christmas and Hanukkah.	[‘Sicilian slice’, ‘pizza’]

3.4.4 Sentiment Analysis

Sentiment analysis is the process of determining the sentiment expressed in a piece of text, such as a review, social media post, or customer feedback. It involves analyzing and classifying the subjective information present in the text to identify the underlying sentiment or attitude conveyed. PyABSA (Yang & Li, 2022) is a pre-trained model we use for the sentiment analysis. Given an aspect, it outputs the probability of the sentiment as follows: negative, neutral, and positive. The probability of each sentiment is between 0 and 1. After retrieving the aspects from the previous step, we then input each review which consists of multiple sentences, and a focal aspect as a pair into the pre-trained PyABSA model and receive the sentiment probabilities. A restaurant may have multiple reviews, and some reviews may discuss the same aspect. The sentiment of the same aspect calculated by averaging the sentiment probabilities obtained from all reviews, so that unique aspects and sentiment scores are used to represent a restaurant. Examples of using the pre-trained model is shown in Table 3-3.

Table 3-3: Example of Sentiment Analysis

ABSA Method	Original Text	Extracted Aspects	Predicted Sentiment
Example 1-LDA	Quite simply, THE go-to restaurant in Clearwater Beach. Try the **grouper sandwich** (Buffalo style) and enjoy their **drink specials**!	burger	pos
		table	neu
		beer	pos
Example 2-Pre-defined	We dared the food even though it was busy. First we had fish tacos, they were good. Those came out of the kitchen within five minutes. Then we ordered the loaded tator tots. They forgot about our order and we had to wait over 45 minutes. Great though, I would order them again.	value for money	neg
		service	neg

doi:10.6342/NTU202303218

	Prices too high. Vodka mixed drink was \$11, double amaretto sours were \$20 each. The beer bucket was \$30 for 6 beers.	food quality	pos
Example 3-InstructABSA	We stopped in for a quick bite while walking the boardwalk and were not disappointed. There is not a ton of seating, so we had a wait a few minutes, but they moved quickly, cleared a table, and got us set up to order. I had the fish tacos, and they were excellent - I would go back to have them again. I also had a taste of the lobster roll and it was amazing - definitely will order that if I'm ever back in town. All in all, this place is somewhat of a hidden gem you don't want to miss!	seating	neg
		table	neu
		lobster roll	pos
		wait	pos

3.4.5 Social-media-based Feature Construction

After obtaining the extracted aspects through multiple channels, we transform the aspects into features for our proposed RESPECT method. We also implement the features that are proposed in Wang et al. (2016) to represent the attractiveness of restaurants, since it is the first to create a score in aspect level to evaluate the compatibility of a restaurant and a location. Both attractiveness features utilize the probability of topics given by the LDA model, which act as aspects for evaluating the restaurant. Continuing along the probabilities gain from the LDA model, the aspect k probability for a restaurant v can be represented as t_{vk}, $\theta_v = \{t_{v1}, t_{v2}, t_{v3}, \ldots, t_{vk}\ |0 \leq\ t_{vk} \leq 1, \text{and } \sum_{k=1}^{K} t_{vk} = 1\}$ is the set of probabilities for all the aspects extracted from LDA. The probability of aspect k at a location is the combination of all restaurants located in location l noted as $A(l)$, which is presented as $t_{lk} = \frac{1}{Z}\sum_{v \epsilon A(l)} \log{(N(v)+1)} t_{vk}$, where $N(v)$ represents the review count of restaurant *v*. With these two features in hand, we can construct the attractiveness features.

Affinity measures the harmony between consumers of restaurant v and those at location l. The formulation is as follow:

$$aff_{vl} = \{t_{vk} * \ t_{lk}, 1 \leq k \leq K\} \tag{3.4}$$

More similar preference of customers in restaurant v and that at location l, the higher the score, and vice versa. If the score is high, the success rate of restaurant v at location l will be higher.

Complementary measures the supplementary between the preference of the consumers of restaurant v and that at location l, which means measuring the opposite preference between them.

$$comp_{vl} = \{t_{vk} * (1 - t_{lk}), 1 \leq k \leq K\} \tag{3.5}$$

The score obtained from analyzing the behavior of the two distinct groups of consumers reveals their contrasting preferences and tendencies, providing valuable insights into their distinct patterns of interaction with businesses and offerings. This disparity in behavior can be viewed as a potential opportunity or a potential threat, contingent on the receptiveness and adaptability of the consumer base at the specific location. Then it comes down to the openness of the consumers at location l. In scenarios where consumers at location l exhibit a sense of openness, curiosity, and a propensity to explore novel experiences and offerings, the observed behavioral differences can be regarded as an opportunity. Conversely, when consumers at location l display a more conservative or resistant attitude towards novel offerings, the identified behavioral contrasts may pose a threat. Failing to align with the prevailing tastes and preferences could result in reduced consumer engagement and limited market acceptance, hindering business growth and expansion in that specific location.

Competitiveness include two features (i.e., strength and opportunity), both composing of two components. First is **importance**. We measure importance by the number of times the aspect was mentioned through all the reviews under discussion. $Importance_{vk}$ refers to the number of times aspect k being mentioned in restaurant

doi:10.6342/NTU202303218

v's reviews and $Importance_{lk}$ refers to the number of times aspect k being mentioned in the reviews of all the restaurants at location l. Second is performance gap. $Performance\_gap_{vl,k}$, just as it sounds, it measures how well the restaurant performs comparing to other restaurants at location l, by subtracting the sentiment score of aspect k under restaurant v's reviews with the sentiment score of aspect k under the reviews from all other restaurants locating at location l.

Strength is a measure by multiplying importance of restaurant v to the performance gap.

$$s_{vl} = \{Importance_{vk} * Performance\_gap_{vl,k}, 1 \leq k \leq K\} \tag{3.6}$$

The higher the score meaning the restaurant outperforms other restaurants at location l on certain aspect that the consumers of restaurant v care about. Since you are satisfying your consumers a lot, it becomes a testament to the restaurant's strength, affirming its ability to meet and even exceed the expectations and preferences of its loyal customer base. These satisfied customers, in turn, contribute to the restaurant's positive reputation, word-of-mouth referrals, and repeat business, further solidifying its position as a preferred choice among the local dining options. In the competitive landscape of the restaurant industry, the restaurant's ability to consistently deliver exceptional experiences and cater to its consumers' desires grants it a competitive edge over rival establishments. This strength positions the restaurant to attract new customers at location l, thereby bolstering its market presence and augmenting its performance.

Opportunity is a measure by multiplying importance of location l to the performance gap.

$$o_{vl} = \{Importance_{lk} * Performance\_gap_{vl,k}, 1 \leq k \leq K\} \tag{3.7}$$

The elevated score demonstrates the restaurant's ability to meet and exceed the expectations of its target customers at location l, thereby fostering a positive reputation and building strong customer loyalty. This competitive advantage positions the restaurant as a preferred choice among the local dining options, attracting a steady stream of patrons and ensuring consistent business success. Given the restaurant's exceptional performance and favorable reception at location l, venturing into opening a new branch or store becomes a logical and strategic move. Capitalizing on the trust and goodwill it has garnered, the restaurant can leverage its successful track record to appeal to a broader customer base and expand its market presence.

Feature selection is implemented on the competitiveness features constructed by pretrained model retrieved aspects, since, through grouping and frequent filtering there are still 180 aspects left. Comparing with the number of aspects generated from former methods, it is significantly larger (almost 5-6 times more). Therefore, we would like to reduce the features to avoid them becoming noises to our model. We implement feature selection through ranking the importance of features created on the basis of these aspects by chi-square test. The chi-square test is especially useful when dealing with categorical variables and a categorical target variable. It helps identify features that have a significant impact on the target variable and can be used to reduce the dimensionality of the dataset, leading to more efficient and potentially more accurate models. We calculate the chi-square score of the competitiveness features between two groups, namely locations that the stores actually open there and locations that stores did not open. Locations are marked as 1 if the store really locates at the place and 0 for the candidates. After the chi-square test, we choose top 25% of the aspect constructed competitiveness feature as our model input.

3.5 Competitor Identification for Social-Media-Based Feature

In our study, we implement competitor identification through community detection. Community detection refers to the process of identifying groups or communities of

doi:10.6342/NTU202303218

entities who exhibit similar patterns. It involves analyzing the relationships or interactions among entities to uncover meaningful clusters or communities. The graph that we use to implement community detection is through the relation of ‘co-mention’. We construct the co-mention matrix with the sampled documents we used earlier for LDA training. The co-mention matrix is undirected, each entry in the matrix corresponds to the count of co-mentions between two restaurants in the reviews. Once the matrix is constructed, we also create a weighted graph using Jaccard similarity coefficient. Then, we use the modularity-based approach to get the best partition and clustering of all the restaurants.

3.5.1 Example of the identified competitor groups

Restaurants are being grouped into 10 clusters. Here are some examples of the groups. After looking through the grouped restaurants in each cluster, we give each cluster a name to represent the cluster (the bolded restaurants are those who are related to the cluster name).

- **Cluster 1** - Mexican Food

  ‘Sauce on the Side’, ‘El Diablo Burritos’, ‘Las Palmas Mexican Restaurant’, ‘Chimi's Fresh-Mex’, ‘California Tortilla’, ‘Grand Hacienda’, ’Bisbee Breakfast Club’, ‘Black Bear Diner’
- **Cluster 2 – Grills and steak**

  ‘Cheddar's Scratch Kitchen’, 'Uno Pizzeria & Grill', ‘Ruth's Chris Steak House’, ‘Rubio's’, ‘Nifty Fifty's’
- **Cluster 3 – American Style**

  ‘Burger Republic’, ‘Dave & Buster's’, ‘Gyro Shack’, ‘Blake's Lotaburger’, ‘Burger Monger’, ‘Village Inn’, ‘Joey's New York Pizza and Italian Restaurant’, ‘SoFresh’, ‘Boise Juice Company’, ‘Token Ramen &

doi:10.6342/NTU202303218

Kung Fu Tea', 'Irish 31', 'Granite City Food & Brewery'

We can see that by creating clusters through co-mentions, these restaurants do display similar categories or offer similar types of meals. Some may seem unrelated to the category type, but it might be that the co-mentions are geographical related meaning that the restaurants are in the same area so it is mentioned, or there are some underlying facts that consumers tend to assume them as competitor or comparative sample.

3.5.2 Feature Construction

Degree of Competitors measures the number of competitors of restaurant v in location l. Restaurants are counted as restaurant v's competitors if they are clustered into the same cluster $G(v)$. Earlier, we have created a Jaccard similarity matrix using the co-mention matrix, we add up the Jaccard weight w_{vi} if the restaurants are in the same group with restaurant v.

$$dc_{vl} = \sum_{i=1}^{V} w_{vi}\,, \text{where } w_{vi} = \begin{cases} w_{vi}, \text{if } i \in G(v) \\ 0 \end{cases} \quad, t \in V \tag{3.8}$$

As the degree of competitors increases, it indicates a higher level of competition in the chosen location, with a greater number of competitors operating at the same area. Value of the degree of competitors reflects a densely saturated market environment, where numerous restaurants battle for the attention and patronage of the local consumer base. This intensified competition can pose both challenges and opportunities for businesses seeking to establish a presence in the area.

Chapter 4 Empirical Evaluation

4.1 Dataset Collection

Our original dataset is obtained from Yelp[7], containing businesses, reviews, and user data for use in personal, educational, and academic purposes. It offers over 6,990,280 reviews, 150,346 businesses, 200,100 pictures and 11 metropolitan areas by 1,987,897 users. Since we only focus on restaurants in our study, business without 'restaurant' or 'food' in their category are excluded. Date are narrowed down to 2017-01-01 ~ 2022-12-31. Restaurants that have 3-14 branches and still open till 2022-12-31 with at least 20 reviews are chosen. The granularity of locations is represented by five-digit postal codes is U.S.. The five-digit postal code is used to identify a specific geographic area, typically representing a city or a part of a city in U.S. For instance, 10001, 10019, 10036 covers Manhattan in New York City; 90210 is Beverly Hills in Los Angeles and 94102, 94103, 94104 belong to San Francisco in California. There are 747 of unique postal codes to choose from in the dataset. By removing areas with less than 5 restaurants opening in that area, a total of 407 postal codes are left. Categories are also provided in business profiles. Each branch has its unique set of categories. Some examples are illustrated in Table 4-1.

[7] Yelp Open Dataset 2023 Available at https://www.yelp.com/dataset (Retrieved on Jan 2023).

doi:10.6342/NTU202303218

Table 4-1: Categories of Different Branches of Restaurants

Restaurant	City	Categories
54th Street Grill & Bar	Arnold	‘Nightlife’, ‘American (Traditional)’, ‘American (New)’, ‘Bars’, ‘Burgers’, ‘Salad’
54th Street Grill & Bar	Riverview	‘Nightlife’, ‘Bars', ‘Beer’, ‘Wine & Spirits’, ‘Mediterranean’, ‘Greek’
Acropolis Greek Taverna	St. Petersburg	‘Greek’, ‘Mediterranean’, ‘Nightlife’, ‘Bars’
Acropolis Greek Taverna	Tampa	‘Greek’, ‘Mediterranean’, ‘Beer’, ‘Wine & Spirits’
Boise Fry Company	Boise	‘American (New)’, ‘Local Flavor’, ‘Fast Food’, ‘Gluten-Free’, ‘Burgers’
Boise Fry Company	Meridian	‘Burgers’, ‘Gluten-Free’, ‘Fast Food’

Table 4-2 shows the statistics of our dataset.

Table 4-2: Statistics of Our Dataset

# restaurants / branches	401 / 1859
# locations	407
# reviews	32,308
Average # of reviews per branch	17.37
# of unique categories	282
Average # of categories per branch	4.81

4.2 Evaluation Design

4.2.1 Dataset Construction

We devise our training and testing set construction process by partitioning the branches under the same restaurant into a 2:1 ratio for each restaurant in our data collection. Following this separation, we pair these branches with negative samples representing locations where the restaurant does not open a store. For each restaurant, a total of 20 locations are carefully selected based on their postal codes (nearby postal codes first). At first, the positive samples are added to the list, then we choose the negative samples by their postal codes that are lower or higher than the average positive sample

doi:10.6342/NTU202303218

postal codes by 500, we add them to list until there are 20 postal unique codes in the list. For each restaurant, we rank the positive samples (i.e., locations that the restaurant opens branches) according to the number of reviews in our dataset.

Regarding the features representing both locations and restaurants, we ensure consistency within each restaurant's group by using the features of the testing group to represent the training group and vice versa. For every location representation concerning a specific restaurant *v*, the location is represented by all other restaurants that operate there, excluding restaurant *v*.

4.2.2 Evaluation Procedure and Metrics

To obtain unbiased and reliable evaluation results, bootstrapping is adopted for every experiment and the score we present is averaged across 10 trails. Below is the metric we use to evaluate the models under examination (including our proposed RESPECT model and the performance benchmarks): precision@1, precision@3, mean reciprocal rank (MRR), and normalized discounted cumulative gain (NDCG).

$$p@k = \frac{\#\ of\ recommended\ items\ @k\ that\ are\ relevant}{\#\ of\ recommended\ items\ @k} \qquad (4.1)$$

$$MRR = \frac{1}{Q}\sum_{i=1}^{|Q|}\frac{1}{rank_i} \qquad (4.2)$$

$$NDCG_k = \frac{DCG_k}{IDCG_k}\,,$$
$$DCG_k = \sum_{i=1}^{k}\frac{rel_i}{log_2(i+1)}, \qquad (4.3)$$
$$IDCG_k = \sum_{i=1}^{|REL_k|}\frac{rel_i}{log_2(i+1)}$$

4.2.3 Training Document Selection

To lower the computation power and time, we randomly picked 1/3 of reviews of every restaurant, to simulate a pool of reviews for all restaurants in our dataset. This pool of reviews is used by LDA and for competitor identification using community detection. For LDA, 20 topics are chosen. We believe that by picking 1/3 of the reviews in the dataset and with the training set consisting of 1/3 of the branches, the chance of peeking through testing branches is relatively small.

4.3 Experimental Setup

4.3.1 Implementation Details

In the context of Latent Dirichlet Allocation (LDA), perplexity and coherence are two commonly used measures to evaluate the quality and performance of a topic model. Perplexity is a measure of how well a trained LDA model predicts a hold-out or unseen dataset. It quantifies how surprised the model is when trying to predict new documents. A lower perplexity score indicates that the model is better at predicting unseen documents. Coherence is a measure of how interpretable and meaningful the topics generated by an LDA model are. It assesses the semantic coherence between words within each topic. Higher coherence values indicate that the topics are more coherent and understandable. Coherence measure considers the pairwise similarity between words based on their co-occurrence patterns within a given corpus. Both measures are used to determine the number of topics modeled by LDA. The pool of sampled documents as previously mentioned is used to perform LDA training and evaluation. Based on the perplexity measure and the coherence measure, the number of topics as 20 seems to be most appropriate (as Figure 4-2 illustrates).

doi:10.6342/NTU202303218

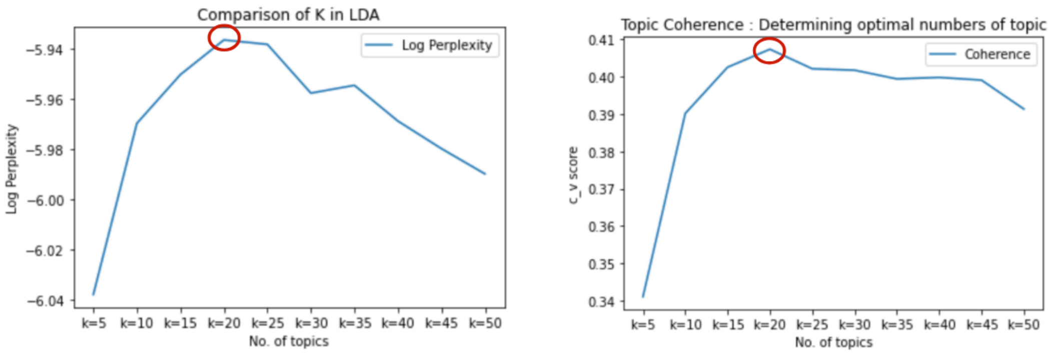


Figure 4-1: Perplexity and Coherence Measures Under Different Numbers of Topics

The Chi-square test of deciding how many aspects should be kept is evaluated through NDCG. We see the performance of the remaining aspects and decide how many percentages of them should be kept. Figure 4-2 shows that retaining 25% of the aspects can achieve the highest score. Hence, for the following experiments, we use the top 25% of the aspects in our RESPECT (InstructABSA) model.

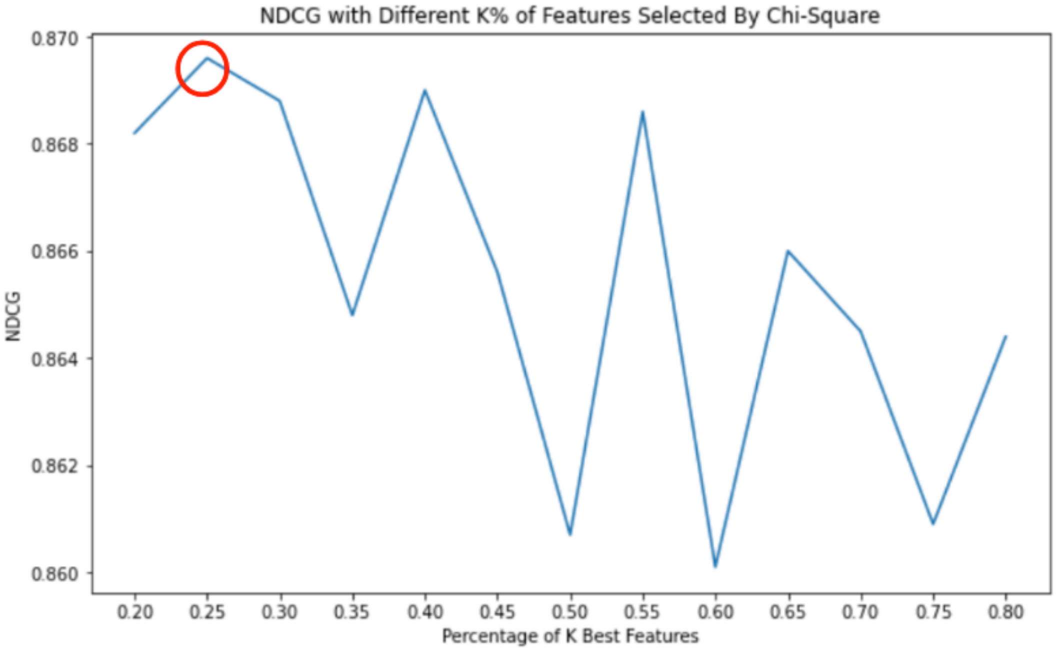


Figure 4-2: NDCG Across Different Percentages of Aspects Selected

As for competitor identification using the modularity approach for community detection, we choose the number of partitions (i.e., clusters) that results in the highest

doi:10.6342/NTU202303218

modularity score. The modularity score represents the quality of a given partitioning of nodes in a graph into distinct communities. It quantifies the degree to which the graph's internal edge density is higher within communities compared to what would be expected in a random graph with the same degree distribution. It measures the difference between the actual number of edges between nodes in the same community and the expected number of edges based on a random graph model. It allows us to evaluate how well a given partitioning of the graph captures the community structure. A higher modularity score indicates a better-defined community structure, where nodes within each community are more densely connected to each other than to nodes in other communities. As Figure 4-3 shows, the best number of partitions (clusters) is 10.

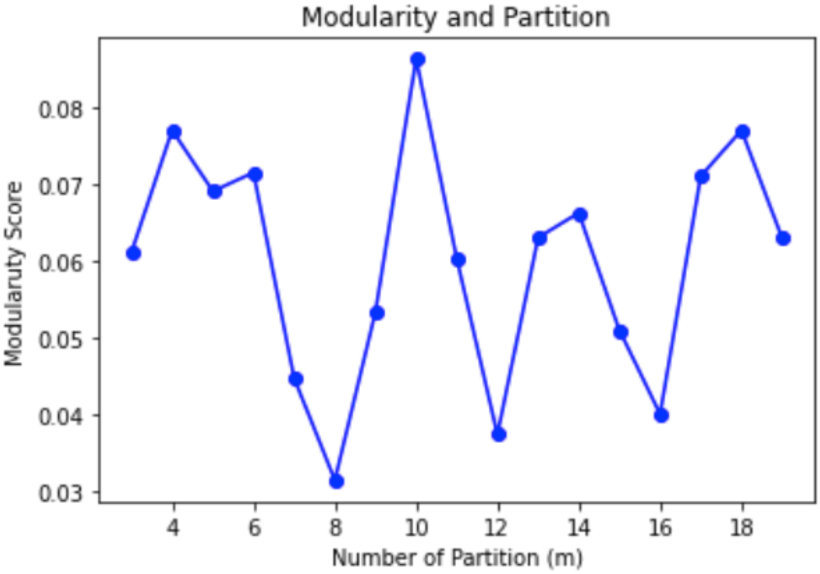


Figure 4-3: Modularity Scores Under Different Number of Partitions (Clusters)

We also adopt and evaluate two ranking methods in this study: pointwise and listwise. The data constructions are as follows. For pointwise ranking, four classification algorithms provided by Scikit-Learn are employed: logistic regression (LR), random forest (RF), support vector classifier (SVC), and decision tree classifier (DTC). As for listwise ranking, LGBMRanker (Ke et al., 2017), a gradient boosting decision tree based

doi:10.6342/NTU202303218

learning-to-rank model, is used. The relevance scores are given by sorting through the number of reviews of the testing branches, starting from 20 and subtracting one for the next lower-ranked branch in sequence.

4.3.2 Comparative Methods

We compare different versions of our proposed RESPECT method a baseline method. The three versions of our proposed method differ in how aspects are extracted: by LDA, pre-defined aspects, and a pretrained model for ABSA (i.e., InstructABSA). As described in Chapter 3, the features involved can be broken down into geo-economics features, location-based features, attractiveness features, competitiveness and degree of competitors features. Specifically,

- Baseline (Wang et al., 2016): Geo-economics features, location-based features, and attractiveness features
- RESPECT (LDA): Geo-economics features, location-based features, attractiveness features, LDA extracted competitiveness and degree of competitors features
- RESPECT (pre-defined): Geo-economics features, location-based features, attractiveness features, pre-defined competitiveness and degree of competitors features
- RESPECT (InstructABSA): Geo-economics features, location-based features, attractiveness features, InstructABSA extracted competitiveness and degree of competitors features

Table 4-3 shows the differences of baseline and our proposed RESPECT method

doi:10.6342/NTU202303218

Table 4-3: Summary of Features Used in Different Methods

Features Used	**Baseline**	**RESPECT (LDA)**	**RESPECT (pre-defined)**	**RESPECT (InstructABSA)**
Geo-economics-based	V	V	V	V
Location-based	V	V	V	V
Attractiveness	V	V	V	V
Competitiveness		V	V	V
Degree of Competitors		V	V	V

4.4 Evaluation Results

Table 4-4 and Table 4-5 show the evaluation results. We can see that three versions of our proposed RESPECT method outperform the baseline method across different ranking algorithms (pointwise ranking and listwise ranking). Pointing out that incorporating sentiment of aspects discussed in reviews do provide useful information since it contains not only preferences towards the restaurant but also how well it performs while ranking with all the other restaurants, making the competitiveness customizes to every restaurant itself. In addition, as mentioned, we identify competitors through the co-mentioning relationships expressed in reviews rather than making use of the categories provided through commercial profiles or tags offered by users. Since co-mentioning in reviews emphasizes the comparison in the consumers' perspective, fusing the viewpoint of consumers into our new store location selection process helps improve the ranking effectiveness.

doi:10.6342/NTU202303218

Table 4-4: Comparison Between Benchmark and RESPECT (Pointwise Ranking)

Method	Model	Precision@1	Precision@3	MRR	NDCG
Baseline (Wang et al., 2016)	LR	6.52%	4.65%	22.71%	35.77%
	SVC	7.76%	4.55%	22.68%	35.23%
	RF	21.57%	14.60%	31.63%	45.60%
	DTC	21.82%	14.60%	31.67%	45.68%
RESPECT (LDA)	LR	6.36%	4.95%	23.14%	36.92%
	SVC	9.66%	7.11%	29.98%	44.28%
	RF	22.45%	17.32%	34.64%	50.00%
	DTC	21.95%	17.37%	37.06%	53.42%
RESPECT (pre-defined)	LR	8.67%	6.11%	23.51%	36.96%
	SVC	12.10%	8.06%	27.48%	40.60%
	RF	27.57%	20.75%	37.80%	54.68%
	DTC	28.82%	24.71%	41.00%	56.74%
RESPECT (InstructABSA)	LR	7.17%	5.02%	24.07%	36.84%
	SVC	9.85%	6.94%	29.75%	43.73%
	RF	22.95%	17.71%	35.05%	50.82%
	DTC	21.95%	17.42%	34.10%	49.97%

Table 4-5: Comparison Between Benchmark and RESPECT (Listwise Ranking)

Method	Model	Precision@1	Precision@3	MRR	NDCG
Baseline (Wang et al., 2016)	LGBMRanker	44.59%	36.31%	59.23%	69.11%
RESPECT (LDA)		58.12%	43.94%	70.20%	76.41%
RESPECT (pre-defined)		50.76%	39.33%	65.03%	73.11%
RESPECT (InstructABSA)		**74.86%**	**53.78%**	**83.18%**	**87.21%**

As Table 4-4 and Table 4-5 also show, the use of listwise ranking algorithm achieves significant higher ranking effectiveness than the use of pointwise ranking algorithm across all methods examined. Moreover, by comparing the three versions of the proposed

doi:10.6342/NTU202303218

RESPECT method, we can see that the use of the pretrained model (InstructABSA) for ABSA outperforms the other two versions. We believe that aspects automatically extracted from reviews may be more down-to-earth, hence, creating more effective competitiveness features for new store location selection than the use of LDA or pre-defined aspects does.

4.5 Additional Evaluation Experiments

4.5.1 Evaluating Features Extracted by Review-based Competitiveness Mining and Feature Selection

In this experiment, our objective is to ascertain the relative contributions of competitiveness features, thereby determining whether it is imperative to incorporate both or if doing so introduces extra noise into our proposed RESPECT model. Specifically, we create different variants of our proposed method: baseline, baseline with degree of competitors, baseline with competitiveness feature, and all features (i.e., the complete RESPECT method). All experiments are conducted on the three versions of the RESPECT method. The experimental results are shown in Table 4-6, Table 4-7, and Table 4-8.

Table 4-6: Result of evaluating features of RESPECT (LDA)

Method	Model	Precision@1	Precision@3	MRR	NDCG
Baseline (Wang et al., 2016)	LGBMRanker	44.59%	36.31%	59.23%	69.11%
RESPECT (LDA) w/o Competitiveness Features		44.66%	36.46%	59.41%	68.99%
RESEPCT (LDA) w/o Degree of Competitors		47.05%	37.37%	61.83%	70.78%
RESPECT (LDA)		58.12%	43.94%	70.20%	76.41%

doi:10.6342/NTU202303218

Table 4-7: Result of evaluating features of RESPECT (pre-defined)

Method	Model	Precision@1	Precision@3	MRR	NDCG
Baseline (Wang et al., 2016)	LGBMRanker	44.59%	36.31%	59.23%	69.11%
RESPECT (pre-defined) w/o Competitiveness Features		44.66%	36.46%	59.41%	68.99%
RESEPCT (pre-defined) w/o Degree of Competitors		54.29%	37.39%	65.92%	71.43%
RESPECT (pre-defined)		50.76%	39.33%	65.03%	73.11%

Table 4-8: Result of Evaluating Features Retrieved by RESPECT (InstructABSA)

Method	Model	Precision@1	Precision@3	MRR	NDCG
Baseline (Wang et al., 2016)	LGBMRanker	44.59%	36.31%	59.23%	69.11%
RESPECT (InstructABSA) w/o Competitiveness Features		44.66%	36.46%	59.41%	68.99%
RESEPCT (InstructABSA) w/o Degree of Competitors		62.18%	48.17%	73.79%	80.03%
RESPECT (InstructABSA)		74.86%	53.78%	83.18%	87.21%

The experimental findings presented in Table 4-6, Table 4-7, and Table 4-8 provide further evidences that the competitive features play a crucial role. The inclusion of these features resulted in a significant enhancement in the evaluation scores, underscoring their utmost importance in our research. Also, the combination of competitiveness features and degree of competitors in some sense show synergy between these two features, because only using the degree of competitors by itself does not make much improvement, but by adding both together makes a considerable progress. Finally, we compare the result before and after using feature selection. As Table 4-9 shows, the ranking effectiveness is significantly improved when using feature selection. The removal of irrelevant aspects greatly enhances the performance of our proposed RESPECT method.

doi:10.6342/NTU202303218

Table 4-9: Result of Evaluating Feature Selection on RESPECT (InstructABSA)

Method	**Model**	**Precision@1**	**Precision@3**	**MRR**	**NDCG**
RESPECT(InstructABSA)	LGBMRanker	67.32%	49.89%	75.88%	82.63%
RESPECT(InstructABSA) with Feature Selection		74.86%	53.78%	83.18%	87.21%

4.5.2 Evaluating Locations with Different Shop Density

Intuitively, restaurants located in popular areas (high restaurant density) are easier to guess. Not only the location presents agglomeration towards consumers acting like a business center luring consumers to flow in, but also contains more information for us to analyze whether the place is suitable for us to open a new branch. As shown in Figure 4-4, we can see that the number of restaurants in every location exhibits a longtail distribution, indicating the imbalance distribution of restaurants across different locations in our dataset.

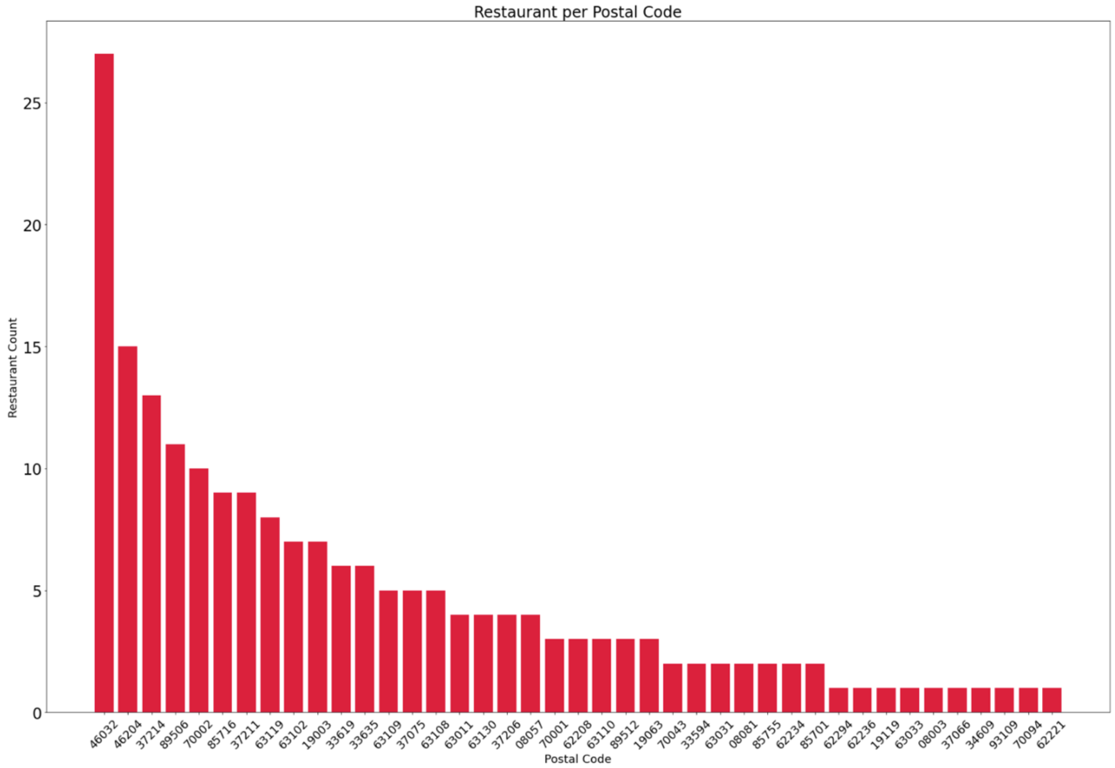


Figure 4-4: Distribution of Restaurants in Different Postal Codes

doi:10.6342/NTU202303218

We use our best performing model–RESPECT (InstructABSA) for this analysis. The evaluation results are in Table 4-10. Predictions for restaurants located in popular areas show promising performance. It is noteworthy that even with a significant disparity of around 10 times in restaurant count between the restaurants in the top third and those in the bottom third, prediction effectiveness remains acceptable. Based on this evaluation results, we can highlight that the benefits of location-based analysis are not limited to urban or popular areas alone. The application of location-based insights extends to rural or less populated regions as well, aiding restaurants in identifying favorable and prosperous locations. In other words, even in rural or less populated areas, our proposed RESPECT method can provide valuable information to assist restaurants in finding optimal sites with good potential for success. This underscores the versatility and effectiveness of review-based strategies in guiding restaurant location selection, regardless of the level of urbanization or popularity of the area.

Table 4-10: Result of Evaluating Locations with Different Shop Density

Locations	Restaurant Count	Model	Precision@1	Precision@3	MRR	NDCG
Top third	305	LGBMRanker	81.07%	56.50%	89.95%	91.04%
Mid third	60		57.50%	39.54%	68.19%	77.85%
Bottom third	36		43.50%	35.67%	57.61%	70.39%

4.5.3 Effects of Percentage of Reviews Used

In our dataset, the review count per restaurant also shows a long-tailed distribution, meaning that the reviews are extremely unbalanced between restaurants. Figure 4-5 demonstrates the distribution of review count per restaurant. This may affect the performance of our proposed RESPECT method. Hence, two experiments are designed

doi:10.6342/NTU202303218

to test if review count has impacts on the effectiveness our proposed method.

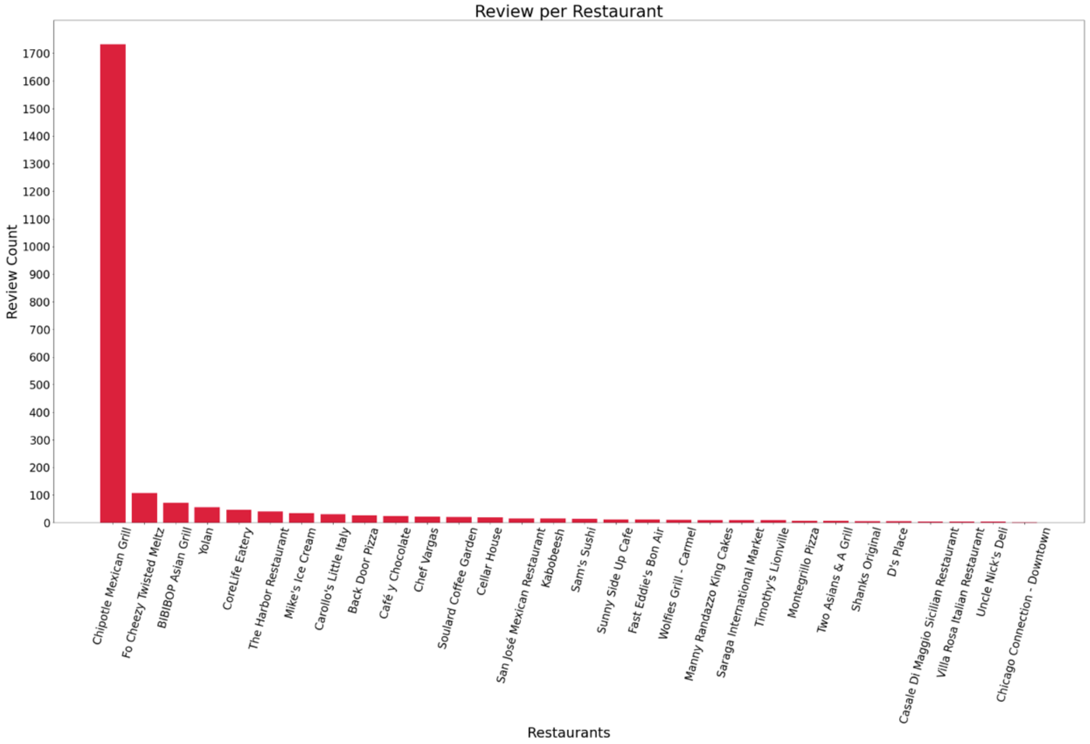


Figure 4-5: Distribution of Reviews Across Different Restaurants

We conduct an experiment to investigate the impact of number of reviews used. To achieve this, we employ different percentages of reviews to represent restaurants in the testing dataset. By conducting this evaluation, we aim to discern the extent of influence that the number of reviews exerts on the overall performance of our proposed RESPECT method. As shown in Table 4-11 we can conclude that our proposed model is not highly sensitive to the number of reviews used. Even when we include only 10% of reviews (for restaurants and locations), our proposed RESPECT method is still capable to make satisfactory recommendations (with precision@1 at 62.20%, precision@3 at 45.16%).

doi:10.6342/NTU202303218

Table 4-11: Result of Evaluating Different Percentage of Reviews Used

Reviews used	**Avg. Review Count**	**Model**	**Precision@1**	**Precision@3**	**MRR**	**NDCG**
100%	80.69	LGBMRanker	74.86%	53.78%	83.18%	87.21%
80%	64.45		68.60%	46.59%	79.20%	85.27%
40%	32.22		64.67%	45.48%	76.55%	83.06%
20%	16.11		63.95%	45.32%	75.67%	82.79%
10%	8.06		62.20%	45.16%	75.32%	81.87%

doi:10.6342/NTU202303218

Chapter 5 Conclusion

5.1 Summary

The key contribution of our research is novel feature provision for the new store location selection task, including ABSA extracted features and **incorporation of community detection for identifying competitors.** This novel set of features specifically designed to aid the task of ranking potential locations for restaurants. ABSA extracted features take into account various factors such as sentiment at the aspect level providing valuable and detailed insights into the suitability of each location. We leverage community detection techniques to identify and analyze competitors in the restaurant industry. Unlike conventional approach that solely relies on direct category matching, our method delves deeper into the network structure and relationships between establishments, revealing hidden patterns of competition. The results of our analysis show that this novel approach outperforms the baseline method, providing valuable insights into the competitive landscape and offering restaurant owners a more comprehensive understanding of their market rivals.

By combining these two contributions, our research delivers a comprehensive and effective framework for new store location selection and competitive analysis. The integration of novel features and community detection not only enhances the accuracy of location ranking but also provides valuable intelligence to guide strategic decision-making for restaurant businesses.

5.2 Future Research Directions

Dealing with word mismatch and disambiguation is worth taking into consideration for future research. In the current study, aspects extracted from InstructABSA are

doi:10.6342/NTU202303218

processed and grouped through heuristic basis, including choosing the aspect groups, classification criteria for fitting words into certain groups. A systematic or rule-based approach can be taken to group these aspects. Co-mentions in reviews have not been disambiguated. Some abbreviations of restaurants may be left out; hence, an incomplete co-mention matrix is constructed and used in our experiment. For example, 'Token Ramen & Kung Fu Tea' may be mentioned as 'Token Ramen' and hence the latter is not considered in the matrix.

Another future research direction is to extend our proposed RESPECT method to handle multi-lingual reviews. Although the reviews we collect from Yelp are all in English, real-world scenarios may encounter to open international branches and in this case we need to deal with multi-lingual reviews. The development of multi-lingual ABSA technique becomes a key.

Cold start problem is also worthy of exploration. As we observe in our experiment, the decrease of review count degrades the recommendation effectiveness, although the performance decline is not significant as compared the decrease of review count. Restaurants with few reviews or few branches may not provide enough information to compute attractiveness and competitiveness features not to mention competitor identification. Finding a way to handle restaurants with few reviews is a path to improve our proposed RESPECT method. To further expand the cold start problem, new open restaurants with no review is also a compelling subject for investigation.

doi:10.6342/NTU202303218

References

Ahedo, V., Santos, J. I., & Galán, J. M. (2021). Knowledge transfer in commercial feature extraction for the retail store location problem. *IEEE Access*, *9*, 132967-132979.

Alharthi, K., El Hindi, K., & Alzahrani, S. M. (2020). Venue-popularity prediction using social data participatory sensing systems and RNNs. *IEEE Access*, *9*, 3140-3154.

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, *114*, 24-31.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993-1022.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, 189-215.

Chen, M., Kuzmin, K., & Szymanski, B. K. (2014). Community detection via maximization of modularity and its variants. *IEEE Transactions on Computational Social Systems*, *1*(1), 46-65.

Damavandi, H., Abdolvand, N., & Karimipour, F. (2019). Utilizing location-based social network data for optimal retail store placement. *Earth Observation and Geomatics Engineering*, *3*(2), 77-91.

Do, H. H., Prasad, P. W., Maag, A., & Alsadoon, A. (2019). Deep learning for aspect-based sentiment analysis: a comparative review. *Expert systems with applications*, *118*, 272-299.

Fox, E. J., Postrel, S., & McLaughlin, A. (2007). The impact of retail location on retailer revenues: An empirical investigation. *Manuscript, Edwin L. Cox School of Business, Southern Methodist University, Dallas, TX.*

Ge, D., Hu, L., Jiang, B., Su, G., & Wu, X. (2019). Intelligent site selection for bricks-and-mortar stores. *Modern Supply Chain Research and Applications*, *1*(1), 88-102.

Guo, B., Li, J., Zheng, V. W., Wang, Z., & Yu, Z. (2018). Citytransfer: Transferring inter-and intra-city knowledge for chain store site recommendation based on multi-source urban data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(4), 1-23.

Hernandez, T., & Bennison, D. (2000). The art and science of retail location decisions. *International Journal of Retail & Distribution Management*, *28*(8), 357-367.

Ismail, N. H., & Hooy, C.-W. (2023). Predicting Restaurant Revenue using Machine Learning. *Journal of Contemporary Issues and Thought*, *13*(2), 9-22.

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull Soc Vaudoise Sci Nat*, *37*, 547-579.

Jensen, P. (2006). Network-based predictions of retail store commercial categories and optimal locations. *Physical Review E*, *74*(3), 035101.

doi:10.6342/NTU202303218

Karamshuk, D., Noulas, A., Scellato, S., Nicosia, V., & Mascolo, C. (2013). Geo-spotting: mining online location-based services for optimal retail store placement. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining,* 793–801

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *30*.

Li, C.-T., Chen, H.-Y., & Zhang, Y. (2021). On exploring feature representation learning of items to forecast their rise and fall in social media. *Journal of Intelligent Information Systems*, *56*, 409-433.

Li, Y., & Liu, L. (2012). Assessing the impact of retail location on store performance: A comparison of Wal-Mart and Kmart stores in Cincinnati. *Applied Geography*, *32*(2), 591-600.

Morrison, P. A., & Abrahamse, A. F. (1996). Applying demographic analysis to store site selection. *Population Research and Policy Review*, *15*, 479-489.

Rahman, M. K., & Nayeem, M. A. (2017). Finding suitable places for live campaigns using location-based services. *Proceedings of the Fourth International ACM Workshop on Managing and Mining Enriched Geo-Spatial Data*, 1–6

Scaria, K., Gupta, H., Sawant, S. A., Mishra, S., & Baral, C. (2023). Instructabsa: Instruction learning for aspect based sentiment analysis. *arXiv preprint arXiv:2302.08624*.

Sokol, V., & Jordanov, K. (2020). Site selection for small retail stores using sustainable and location-driven indicators: Case study: Starbucks coffee shops in Los Angeles. Unpublished Master Thesis. Department of Industrial Economics, Blekinge Institute of Technology, Karlskrona, Sweden.

Song, Y.-Y., & Ying, L. (2015). Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, *27*(2), 130.

Taneja, S. (1999). Technology moves in. *Chain Store Age*, *75*(5), 136-138.

Wang, F., Chen, L., & Pan, W. (2016). Where to place your next restaurant? Optimal restaurant placement via leveraging user-generated reviews. *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 2371–2376

Wright, R. E. (1995). Logistic regression.

Xu, M., Wang, T., Wu, Z., Zhou, J., Li, J., & Wu, H. (2016). Demand driven store site selection via multiple spatial-temporal data. *Proceedings of the 24th acm sigspatial international conference on advances in geographic information systems,* 40, 1–10

Yang, H., & Li, K. (2022). PyABSA: open framework for aspect-based sentiment analysis. *arXiv preprint arXiv:2208.01368*.

Yu, Z., Tian, M., Wang, Z., Guo, B., & Mei, T. (2016). Shop-type recommendation leveraging the data from social media and location-based services. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *11*(1), 1-21.

Zeng, J., & Tang, B. (2019). Mining heterogeneous urban data for retail store placement. *Proceedings of the ACM Turing Celebration Conference-China,* 53, 1–5

Zhdanov, P., Glukhov, G., & Derevitskii, I. (2022). Machine Learning Approach to Optimal Restaurant Location Search. *Proceedings of IEEE 16th International Conference on Semantic Computing (ICSC)*, 263-268

doi:10.6342/NTU202303218