2021 4th International Workshop on Advances in Social Sciences (IWASS 2021)

Local Store Valuation: A Newly-established Concept Developed from Site-Selection Problem

Zhiying Chen[1], Ziye Zhou[1], XuYUAN[2]

HJYCloud Technology, School of Software, Dalian University of Technology, Shenzhen, Guangdong, China

Keyword: Optimal, Location, Valuation, Ranking, Algorithm.

Abstract: Finding the best site for a local store can be defined as a site-selection problem. Site-selection is a strategic decision that involves several limited criteria with consideration for technical, economic, environmental, and demographic perspectives. The appropriate candidates for site-selection are usually fixed to a few locations. In this paper, we created a brand-new concept, local store valuation, developed furtherly from site-selection problem, specifically, the objective of this problem not only decides for finding the most suitable location, but also provides a score or a ranking for each potential location. Additionally, we initialized a new approach for this problem with the combination of the main methods in the site-selection problem. Manufacture and sales are beneficial from this new approach.

1. Introduction

A problem related to store's location with a long history in research is called site-selection problem. A site-selection problem for different kinds of organizations could be focused on a wide range of industries, plant location selection, store and warehouse location selection, healthcare and hospital location selection [1].

A new problem, extended from this traditional site-selection problem, named local store valuation, is mainly focused on offering respective quantitative indicators by considering technical, economic, environmental, and demographic factors when comparing the whole alternatives for site location. Furthermore, these quantitative indicators are ultimately converted into a single score rather than presenting multi-faceted assessments in related fields as in the traditional site-selection problem. This new problem features in offering a score for both potential candidates and existing locations rather than potential candidates. There exist two minor differences as well. The first one is that the potential locations for a retail store are no longer limited in our newly-defined problem. More formally, the representation for a location is the coordinate of latitude and longitude rather than a single site without quantitative measurement. Secondly, the method to this new problem considers all fields of aforementioned factors instead of one or two kinds of factors involved in the previous site-selection problem.

2. Related Study

Models for such a site-selection problem in retail stores are mainly divided into two streams, empirical models and machine learning algorithms. Empirical models dominated the research method in this problem before the prevalence of wide application in machine learning algorithms and location-based social networks (LBSNs) [2]. Gravity model [3], Huff model [4], and Analytical Hierarchy Process (AHP) [5] are the typical representations among empirical methods. Machine learning techniques such as regression models [6], linear regression [7], support vector machine [7], and decision trees [7] were applied to site selection.

According to the research done by Porta et al. [8], and Wang et al. [9], street centrality imposes a huge impact in the location of local retail stores and services. They follow the similar pipeline that not only applied Multiple Centrality Assessment (MCA) including closeness, betweenness and

  DOI: 10.25236/iwass.2021.040

straightness to estimate the centrality, but also introduced Kernel Density Estimation for the measurement of the relationship between street centrality and spatial distribution for different local stores and services.

Global Positioning System (GPS) and Wi-Fi data, i.e., social media's check-ins, location-sharing, act as a catalyst for researchers to understand the public engagement and volume of people, thereby referring as a measurement of how many people potentially walk into a local store [10]. However, there are only a few related research [2] either in this field. This paper will continue the exploration in this direction as including LBSNs data as a quantitative measurement.

Traffic accessibility measuring travel time and distance for customers to reach a local store was merely mentioned in the previous literature except for the Huff model and machine learning algorithms applied recently [11] which depends heavily on the calculation of distance. Although Huff model and its variants have been proved to be powerful to guide a new store location placement, it is still limited due to its idealization and the factors they considering.

Wang, et al. [10] and Yang, et al. [7] analyzed the site selection for digital signage and hotel based on the introduction of backpropagation neural network and various machine learning models respectively. These two methods paved the way for our method to apply machine learning on a retail store with various commodities, rather than a single industry.

3. Method

The influential factors are concluded and summarized to three main categories, street centrality, public engagement and travel cost. The quantitative measurements for these three perspectives are either improved from a previous related study or integrated with several existing methods.

3.1 Street Centrality

MCA measures the centrality of street networks, for instance, straightness, closeness, and betweenness.

The betweenness centrality depends on the percentage of number of the shortest paths connecting two arbitrary nodes passing a specific node of the total number of the shortest paths.

$$\mathrm{B}(x) = \sum_{y,z} \frac{n(y,z)^x}{n(y,z)}$$

where $n(y,z)^x$ denotes the number of the shortest paths connecting node y and z traversing x and $n(y,z)$ denotes the number of the shortest paths between node y and z.

The closeness centrality measures the average distance of the shortest paths from a specific node to an arbitrary node.

$$\mathrm{C}(x) = 1/\sum_{y} d(y,x)$$

where $d(x,y)$ indicates the shortest distance between node x and y.

3.2. Public Engagement

The measurement of street centrality contains no information about flow of customer traffic. Street centrality does not consider the effect of customer traffic. The combination of street centrality and public engagement provide a more comprehensive landscape for location valuation. In addition, the rapid development of LBSNs data alleviates the difficulty of measuring public engagement via social media check-ins data and location-sharing.

In order to measure the traffic flow of customers for an arbitrary location, we count the total number of media check-ins and location sharing empirically collected among immediate surroundings from this location in a fixed threshold t as follows:

$$T_{actual}(x|t) = |\{(x,l) \in D_C : l \leq t\}|$$

where tuple (x, l) denotes check-in recorded in place x within immediate surroundings from x in a l -distance, and D_C is a dataset of all check-ins recorded.

3.3. Traffic Accessibility

Traffic accessibility denotes the cost for a travel in terms of linear distances and actual distance traveled along the street, and time consumed. Traffic accessibility can be represented as follows:

$$A_{d-linear} = d(x, y) \quad A_{d-linear} = d(x, y|x \to y) \quad A_t = t(x, y|x \to y)$$

where $A_{d-linear}$ denotes traffic accessibility measured by linear distance between x and y, $A_{d-actual}$ is the actual distance, and A_t is the time consumed when transiting from x to y. $x \to y$ represents the ride from x to y along the street.

The main difference of our measurement for traffic accessibility compared to previous studies is that, the objective of this indicator is to define an approximate demand for an arbitrary community based on the GPS origin-destination track data. Assuming that C is a community, OD_c is a dataset of all origin-destination track records started from C and P is the target location we are measuring, we have that

$$T_{potential}(P|C) = |\{(p, c) \in OD_c : C \to P\}|$$

where $T_{potential}(P|C)$ is the potential demand for location P in community C, and $C \to P$ denotes the origin-destination pair is $C - P$. P is defined as one of the most frequently visited destinations among the OD.

Hence, public engagement measures the actual volume of customers, by contrast, traffic accessibility not only denotes the accessibility for a location, but also entails the information about potential customers.

3.4 RankNet

We consider a pair-wise learning-to-rank approach, RankNet [7] which learns the ranking for all possible and existing locations based on the features aforementioned. The architecture of our proposed RankNet is shown as follows:

The base network with the traditional softmax function removed from the end. The inputs of this network are feature embeddings for two locations and the outputs are two scoring functions. These scoring functions are inputs for the ranking network which uses binary cross-entropy loss function to train. The ranking network firstly take the difference between the scoring functions and then pass it to a sigmoid activation function.

This network is able to compare locations given the characterized features in a pairwise manner. The final output of this RankNet is either a ranking or a scoring, depending on the requirement of the local store valuation.

3.5 Algorithms

After the collection of multiple features and the selection of machine learning algorithms, we propose our algorithm as follows:

1. Evaluating for street centrality, public engagement, and traffic accessibility.
2. Training the machine learning algorithms for selected business success indicator.
3. Applying RankNet based on all the features return a ranking.

This general framework provides a flexible direction for the valuation of different industries. Minor adjustments are easily to implement according to the various purposes for the local retailers.

4. Conclusion

This paper defines a new problem named local store valuation which is developed from the traditional site-selection problem for a local store. The new problem aims to offer a score or a ranking for not only potential candidates when deciding a new site but also the existing stores according to multi-faceted factors from technical, economic, environmental and demographic perspectives.

The general framework of valuation algorithm introduced by this paper can be theoretically generalized to all walks of life since the majority of influential factors to capture almost all potential linear and non-linear relations. To the best of my knowledge, this is the first comprehensive algorithm for local store valuation, which is the main contribution to the related research followed the newly-defined problem.

The valuation learned from algorithm guides retailers to better contribute their products, supply chains, logistics and capitals. Moreover, algorithm provides a quantitative valuation for each store based on more than economic aspect. Equipped with the traditional business success indicators, this valuation offers much more meaningful insights for the rationale behind consumption and management loans set by bank or financial corporation.

This research has a few limitations as it does not consider the effect of potential competitiveness for the similar type of nearby stores and future development of the area located by the store. We will further explore more in a spatial-temporal model to capture both spatial and temporal effects for a local store.

Reference

[1]. T. Hernandez, "Enhancing Retail Location Decision Support: The Development and Application of Geovisualization", *Journal of Retailing and Consumer Services,* 14, 2007, pp.249-258.

[2]. Yu et al., "Shop-Type Recommendation Leveraging the Data from Social Media and Location-Based Services". *ACM Transactions on Knowledge Discovery from Data,* vol.11, no.1, 2016, pp.1-21.

[3]. J. E. Anderson, "The Gravity Model", *Annual Review of Economics,* vol.3, 2011, pp.133-160.

[4]. D. L. Huff, "A Probabilistic Analysis of Shopping Center Trade Areas." *Land Economics,* vol.39, no.1, 1963, pp.81-90.

[5]. R. Handfield et al., "Applying Environmental Criteria to Supplier Assessment: A Study in the Application of the Analytical Hierarchy Process", *European Journal of Operational Research,* vol.141, no.1, 2002, pp.70-87.

[6]. M. Xu, et al., "Demand Driven Store Site Selection via Multiple Spatial-Temporal Data", presented at Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2016, pp.1-10.

[7]. Y. Yang, et al., "Hotel Location Evaluation: A Combination of Machine Learning Tools and Web GIS.", *International Journal of Hospitality Management,* vol.47, pp.14-24.

[8]. S. Porta, et al., "Street Centrality and the Location of Economics Activities in Barcelona.", *Urban Studies,* vol.49, no.7, pp1471-1488.

[9]. F. Wang, et al., "Location Analysis of Retail Stores in Changchun, China: A Street Centrality Perspective.", *Cities,* vol.41, 2014, pp.54-63.

[10]. Y. Wang., et al., "Site Selection of Digital Signage in Beijing: A Combination of Machine Learning and an Empirical Approach.", *ISPRS International Journal of Geo-Information,* vol.9, no.4, 2020, pp.217.

[11]. M. Xu, et al., "Store Location Selection via Mining Search Query Logs of Baidu Maps". *Computing Research Repository,* vol.1606, no.4, 2016, pp.17-28.