Utilizing Skip-Gram for Restaurant Vector Creation and Its Application in the Selection of Ideal Restaurant Locations

Chih-Yung Chang(✉), Syu-Jhih Jhang, Yu-Ting Yang, Hsiang-Chuan Chang, and Yun-Jui Chang

Tamkang University, New Taipei City, Taiwan
cychang@mail.tku.edu.tw, {812410016,149190}@o365.tku.edu.tw, 142720@mail.fju.edu.tw

Abstract. Restaurant Site Selection (RSS) plays a pivotal role in the success of launching a new restaurant. The core elements of RSS encompass foot traffic and the consumption capacity potential at prospective sites. Previous studies often relied on data gleaned from social media or the Internet, utilizing statistical or machine learning methods to predict foot traffic. Nevertheless, amassing comprehensive data on foot traffic and consumption capacity proves arduous. Multiple factors, such as MRT flow, bus traffic, and business districts, contribute to foot traffic, rendering data collection complex. Similarly, quantifying consumption capacity involves variables like salary and the habits of residents and workers in the vicinity, posing data collection challenges. In contrast to prior work, this study derives proximity insights from numerous restaurant types and their locations. Employing the n-skip gram mechanism from natural language processing, restaurant vectors are generated for each restaurant type. These vectors subtly encapsulate information about foot traffic and consumption capacity. Subsequently, the algorithm utilizes these Restaurant Vectors to recommend optimal restaurant locations. Performance assessments confirm that the generated Restaurant Vectors effectively encompass features related to foot traffic and consumption capacity.

Keywords: *n*-skip gram · neural network · restaurant vector · people flow · consumption capacity · restaurant site selection

1 Introduction

The modern job market is witnessing a shift, with more individuals opting for entrepreneurship, particularly in the restaurant industry. This trend underscores the importance of meticulous planning and evaluation, with the choice of restaurant location taking center stage as a critical initial step.

Technology has revolutionized this process. In the past, information was gleaned from traditional media sources, with limited interaction between consumers and businesses. However, the advent of the internet, coupled with 5G location-aware services,

D.-J. Deng and J.-C. Chen (Eds.): SGIoT 2023, LNICST 557, pp. 141–147, 2024.
https://doi.org/10.1007/978-3-031-55976-1_14

has transformed the landscape. People now access the internet effortlessly through various devices, leaving digital footprints that generate a wealth of data. This social data offers invaluable insights for restaurant owners, aiding in site selection and enhancing operational efficiency.

Traditionally, site selection relied on manual data collection and statistical methods. However, artificial intelligence (AI) has emerged as a powerful tool for collecting and analyzing vast online data. Researchers have begun harnessing AI techniques, including machine and deep learning, to predict optimal restaurant locations with greater accuracy.

This paper builds upon previous AI methods, aiming to address existing challenges. It recognizes that restaurants on the same street share similar features, such as foot traffic and consumption capacity. Leveraging neighbor relations, restaurant vectors are constructed and used to identify the best street for a new restaurant. This innovative approach enhances the accuracy of site selection, offering valuable contributions through the algorithm.

1. ***Restaurant Vector Construction:*** This paper creates restaurant vectors to capture semantic neighboring relations between restaurants, encompassing foot traffic and consumption capacity.
2. ***Data Collection Challenges Overcome:*** The study sidesteps data collection difficulties regarding foot traffic and consumption capacity by relying on restaurant relationships instead.
3. ***Restaurant Vector Validation:*** The study validates restaurant vectors by confirming alignment with the average-person concept regarding restaurant location.

2 Related Work

Numerous studies have aimed to identify ideal business locations, utilizing various data sources including geography, movement, social data, and online government records. Some explored common factors affecting business success [2], while others incorporated machine learning techniques with carefully chosen features or advanced deep learning methods.

Researchers have expanded their focus beyond economic factors, realizing they are not the sole determinants of sound business location decisions. While early studies emphasized cost minimization and profit maximization, globalization and regulatory changes have highlighted the importance of noneconomic factors such as place image, brand, visual appeal, reputation, sense of place, and identity in decision-making. These factors, along with economic data, have become valuable inputs for recent advances in machine learning and deep learning [3].

While machine learning has its merits, it sometimes requires expert knowledge of predetermined features [7, 8, 10–12]. Consequently, some researchers turn to deep learning, which excels with abundant input data. Data sources like user-generated reviews from platforms like Yelp [1], movement data, and geographic features offer rich data for deep learning models. These approaches often outperform traditional proximity-based methods and contribute to practical tools for selecting optimal locations, as demonstrated in this paper.

3 The Proposed Model

To address previous technological limitations, this study introduces a shop location selection system consisting of a store word vector model, a street database, and a processor. The street database contains various street data, each with multiple store data entries, each specifying a store category. The store word vector model associates a word vector with each store category. The processor functions to evaluate recommendation scores for each street data based on the store categories in the street database and a target category. Subsequently, it provides location recommendations based on these scores.

The study also presents a store word vector model training method, involving the processor categorizing store data in multiple street data entries to derive various store categories. Then, the processor trains word vectors for each street data entry based on the store categories within it, establishing the store word vector model. This approach leverages the concept of word vectors, where the similarity between store categories implies a higher likelihood of co-location in the same street. Thus, users, when seeking a location for opening a shop, can set a target category. The shop location selection system utilizes the target category, the store word vector model, and the street database to suggest suitable streets for the target category, offering users valuable guidance for their shop location decisions.

To address this, our research presents a location selection system comprising a Store Word Vector Model, a Street Database, and a Processor. The Street Database holds data from various streets, with each street containing store data categorized by store type. The Store Word Vector Model generates word vectors based on store categories. The Processor utilizes this data to determine recommendation scores for each street based on the store categories in the Street Database and a target category. It then offers location suggestions. Our approach starts by collecting store data from multiple streets, obtained from sources like government records, public data, and Google Maps. This data serves as a reference for store categories, encompassing store names, product information, and services offered. The Processor classifies store data within street entries to derive diverse store categories. Semantic analysis tools, such as Natural Language Processing (NLP), may assist in understanding products or services based on store names. Classification can also consider factors like product types or service categories. In cases where store names are ambiguous, predefined data or manual classification can be applied. In some scenarios, the analysis focuses solely on stores within the same industry, such as food services or clothing. For instance, when used for restaurant site selection, only food-related businesses within street data are categorized based on their product types.

Furthermore, store classification may extend to multiple industries within the same street data. This approach enables word vector analysis across different categories. Additionally, stores can be categorized by price variations for the same products or services. For example, high-end and budget-friendly versions of the same store category can be differentiated, offering granularity in categorization. After classification, each store category is encoded numerically for subsequent processing, making use of methods like one-hot encoding. This approach streamlines location selection by leveraging word vectors to assess the suitability of streets for a target category, enhancing the decision-making process.

The processor trains restaurant word vectors based on restaurant categories within the same street data. Word vectors represent word relationships in a vector space, where closer vectors indicate frequent co-occurrence. Similarly, when restaurant categories often coexist on a street, their word vectors become close.

Various methods like BERT, GPT, ELMO, or word2vec are used to train these vectors. In some cases, the processor selects one category from the same street and another from the opposite side for training. For instance, if there are three adjacent categories like tea shops, bento shops, and dessert shops, training can involve bento shops with tea shops and dessert shops.

The processor calculates the similarity between the target category and restaurant categories. The results are summed for each street, producing recommendation scores. These scores rank streets for potential restaurant locations, helping users make site selection decisions.

The following formula (1) is used to calculate the association between the target category and the restaurant category, where π_i represents the vector of the target category, and π_j represents the vector of the restaurant category. The result of this formula falls between –1 and 1. A value of 1 indicates a high association between the target category and the restaurant category when the two vectors have the same direction. Conversely, a value of –1 suggests a very low association when the vectors point in opposite directions.

$$\mathrm{Sim}\left(\pi_i, \pi_j\right) = \frac{\pi_i . \pi_j}{\|\pi_i\| \left\|\pi_j\right\|} \tag{1}$$

In Fig. 1, let's delve into our proposed *Restaurant Vector Creation* (*RVC*) algorithm, guided by the aforementioned design principles. Initially, s_i represents a certain street. Each street $s_i = \{a_{i,1}, a_{i,2} \ldots, a_{i,|s_i|}\}$ is composed of several existing restaurants $a_{i,j}$. Each s_i is evaluated to calculate its $score_i$, signifying its suitability as the target location. In step 1, each street s_i in the set S undergoes steps 2 to 12. In step 2, each street s_i has initial value $score_i = 0$.

Given a restaurant a_{target}, step 3 aims to develop a location selection mechanism, which determines the location of the restaurant a_{target} to open a shop. Consider each restaurant $a_{i,j} \in S$ and execute the operations outlined in steps 5 to 12. In steps 5 to 7, street s_i earns points if A_k is not equal to A_{target}, indicating that $a_{i,j}$ and a_{target} belong to different restaurant categories. The points awarded are determined by the similarity of the restaurant vectors between $a_{i,j}$ and a_{target}, denoted as $\mathrm{Sim}(\pi_{target}, \pi_k)$. Conversely, if A_k equals A_{target} ($A_k = A_{target}$), street s_i loses points due to the competition between similar restaurants within the same street.

Steps 8 to 12 correspond to the above-mentioned operations. To ensure fairness given varying numbers of restaurants on different streets, step 10 normalizes the $score_i$ by the number of restaurants on the street s_i. Ultimately, the street with the highest score is recommended as the optimal location for the restaurant a_{target}.

Algorithm: ***Restaurant-Vector Creation(RVC)***

Input: The target street a_{target}

Output: The best location s_{best} in region R.

```
for each s_i in S:
  { score_i = 0;
    n_i = 0;
   for each a_i,j in s_i:
    {let A_k = g(a_i,j);
     if A_k ≠ A_tcrget:
         temp= Sim(π_target, π_k)
         if temp≥ λ:
             score_i=score_i+temp
             n_i + +;
       }
   score_i=(score_i/ n_i) }
s_best =arg max_{s_i∈S} score_i
```

Fig. 1. The *RVC* Street Recommend Algorithm.

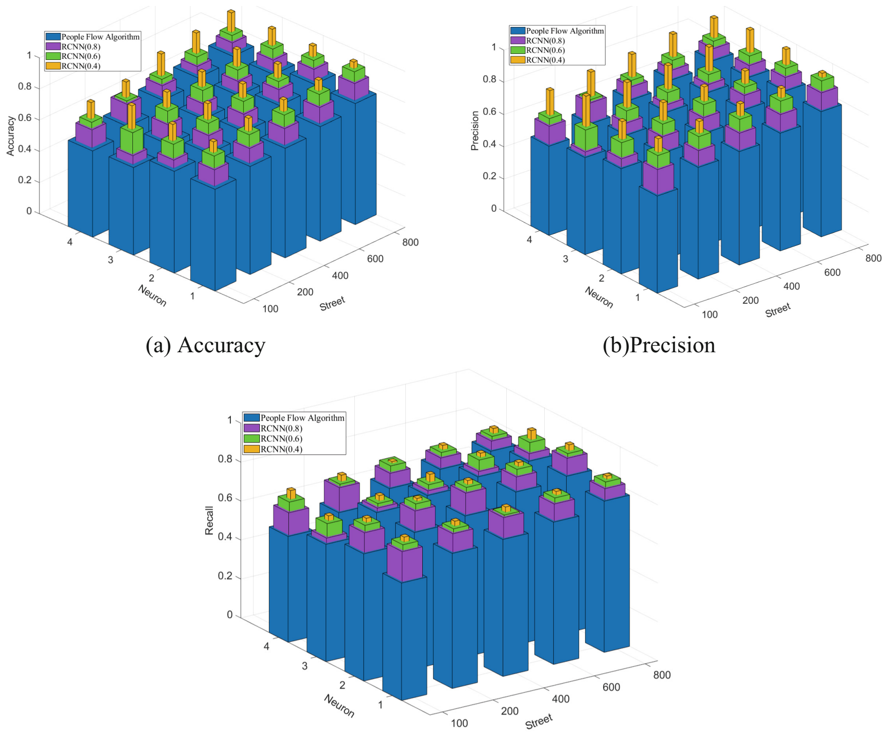
4 Performance Evaluation

This section evaluates the performance of the *RVC*(Restaurant Vector Creation) algorithm in comparison to the People Flow algorithm for restaurant site selection. The People Flow algorithm relies on data like MRT stations, bus stops, business districts, schools, office buildings, and parks to identify suitable restaurant locations. In contrast, RVC uses an RCNN network to assign each restaurant category a vector, trained based on the relations between neighboring restaurants on the same street. The similarity between these vectors determines the frequency of neighboring restaurants.

Figure 2 primarily compares the performance of the proposed RCNN with the existing People Flow Algorithm, focusing on accuracy, prediction, and recall metrics. RCNN(λ) denotes the RCNN performance achieved by setting a threshold value λ, with RCNN(0.8), RCNN(0.6), and RCNN(0.4) representing performances using λ values of 0.8, 0.6, and 0.4, respectively. The People Flow Algorithm recommends streets based on collected information related to People Flow, such as the count of individuals entering/exiting MRT and bus stations, and the number of convenience stores, among other factors.

The experiments vary the number of neurons from 1 to 4 and the number of streets from 100 to 800. Results consistently show that the performance of the three RCNN mechanisms improves with the increasing number of streets in terms of accuracy, precision, and recall. This improvement is attributed to the larger dataset, allowing the RCNN network to extract better relations from neighboring restaurants, resulting in enhanced accuracy, precision, and recall. Conversely, the accuracy, precision, and recall of the existing People Flow algorithm remain constant and are lower than those of the proposed *RVC*(*Restaurant Vector Creation*) algorithm.

Furthermore, the three RCNN mechanisms achieve the highest accuracy, prediction, and recall when the number of neurons is set to two. This is because increasing the number



(a) Accuracy (b)Precision

(c)Recall

Fig. 2. Performance of two compared algorithms and three weights of *RVC* in terms of accuracy, precision and recall.

of neurons spreads the values of the restaurant vectors across more dimensions, reducing cosine similarity comparisons and, consequently, accuracy, precision, and recall.

In contrast, the proposed *RVC* algorithm outperforms the traditional People Flow algorithm in all scenarios. This superiority stems from the People Flow algorithm's inability to collect complete information affecting accuracy, precision, and recall, particularly regarding consumption-related factors like housing prices and people's salaries. In contrast, RVC leverages relations within neighboring restaurant information and represents them using restaurant vectors. These vectors encompass all features impacting restaurant openings, resulting in the superior performance of *RVC* over the People Flow Algorithm.

5 Conclusions

This paper delves into the relationships between neighboring restaurants to create restaurant vectors, which carry the semantic meaning that closer vectors of two restaurant categories suggest they are frequently found on the same street. Restaurants located on the same street often share common characteristics, including foot traffic and consumption capacity. Utilizing restaurant vectors to identify the best-suited street, with features similar to the target restaurant, can encompass all the implicit attributes that influence the restaurant business. Instead of the arduous task of collecting comprehensive data on foot traffic and consumption capacity, the restaurant vector offers a solution.

To our knowledge, the concept of a restaurant vector signifies the similarity of the target restaurant. The novel aspect of our proposed *RVC*(*Restaurant Vector Creation*) algorithm lies in its ability to generate restaurant vectors for each restaurant category and employ them to recommend suitable locations for opening new restaurants. We believe that restaurant vectors hold the potential for broader applications in the restaurant industry. Experimental results affirm that *RVC* surpasses traditional algorithms that rely on data collected from social media or the internet and predict foot traffic using statistical or machine learning methods.

References

1. Tayeen, A.S., Mtibaa, A., Misra, S.: Location, location, location! In: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (2019)
2. Dixit, A., Clouse, C., Turken, N.: Strategic business location decisions: importance of economic factors and place image. Rutgers Bus. Rev. **4** (2019)
3. Bilen, T., .Erel-Ozcevik, M., Yaslan, Y., Oktug, S.F.: A smart city application: business location estimator using machine learning techniques. In: IEEE International Conference on High-Performance Computing and Communications (2018)
4. Bhole, J., Nandiyawar, S., Pawar, S., Vora, P.: Smart site selection using machine learning. Int. Res. J. Eng. Technol. **7**(5), 3012–3015 (2020)
5. .Mazhi, K.Z., Suryana, L.E., Davi, A., Dewi, W.R.: Site selection of retail shop based on spatial analysis and machine learning. In: International Conference on Advanced Computer Science and Information Systems (ICACSIS) (2020)
6. Han, S., Jia, X., Chen, X., Gupta, S., Kumar, A., Lin, Z.: Search well and be wise: a machine learning approach to search for a profitable location. J. Bus. Res. **144**, 416–427 (2022)
7. Eravci, W.R., Bulut, N., Etemoglu, C., Ferhatosmanoglu, H.: Location recommendations for new businesses using check-in data. In: IEEE 16th International Conference on Data Mining Workshops (ICDMW), pp. 1110–1117 (2016)
8. Chang, T.-H.: Restaurant location selection by utilizing the fuzzy preference relations. In: IEEE International Conference on Industrial Engineering and Engineering Management (2010)
9. Wang, Y., Li, S., Zhang, X., Jiang, D., Hao, M., Zhou, R.: Site selection of digital signage in Beijing: a combination of machine learning and an empirical approach. Int. J. Geo-Inf. **9**(4), 3012–3015 (2020)
10. Furtado, A.S., Fileto, R., Renso, C.: Assessing the attractiveness of places with movement data. J. Inf. Data Manag. **4**, 124–133 (2013)
11. Quan, X., Wenyin, L., Dou, W., Xiong, H., Ge, Y.: Link graph analysis for business site selection. Computer **45**(3), 64–69 (2012)
12. Wang, F., Chen, L., Pan, W.: Where to place your next restaurant? Optimal restaurant placement via leveraging user-generated reviews. In: The 25th ACM International on Conference on Information and Knowledge Management (2016)