Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Optimizing Commercial Site Selection Using the Thompson Sampling Algorithm

Hanyue Liao*
College of Artificial Intelligence, Xi'an JiaoTong University, Xi'an, China
* Corresponding Author Email: alvit-frey0213@stu.xjtu.edu.cn

Abstract. In the era of big data, businesses are inundated with vast amounts of information across a wide spectrum of types, creating both opportunities and challenges in data utilization. Among the various algorithms adapted for this environment, the Thompson Sampling algorithm stands out due to its exceptional capability to handle dynamic data and complex problems. This algorithm leverages probabilistic models to continuously update and optimize decision-making processes based on incoming data streams. Thompson Sampling excels particularly in determining the optimal locations for businesses by analyzing extensive databases integrated with sophisticated business location models. By assessing and adjusting to real-time data, it allows for a more nuanced understanding of geographical and demographic trends that are crucial for strategic business placement. This method not only enhances the accuracy of business decisions but also significantly improves adaptability in rapidly changing market conditions.

Keywords: Thompson Sampling Algorithm; Commercial site selection; Multi-armed Bandit.

1. Introduction

Commercial location selection is crucial in business development and management, and is widely acknowledged as a key strategic decision [1]. In determining optimal business sites, factors such as labor costs, aesthetic considerations, and transportation accessibility are crucial for covering expenses and generating profit, thus enabling companies to operate effectively. For retail stores and restaurants, traffic volume, population density, and consumer demand—increasingly influenced by social media—are critical metrics in site evaluation. Conversely, for warehouses and distribution centers, the primary considerations are transport capacity and logistical convenience [2]. With the integration of social media influence and geographic location as evaluation indices, the site selection process also takes competitors and local instances into account. This approach employs decision trees and pruning algorithms to refine choices. Given the growing complexity of these environments, algorithms such as the Shuffled Frog Leaping Algorithm and genetic algorithms have been deployed to tackle NP-hard problems [3]. Additionally, given the inherent uncertainties in site selection, the Analytical Hierarchy Process (AHP) combined with Geographic Information Systems (GIS) has been established as a highly effective method [4].

However, many of the aforementioned approaches either rely on partial factors or tend to converge at local optima. Methods that aim to maximize profit through complex algorithms or extensive databases often fall short in balancing quality and cost effectively. The advancement in computer simulations of human cognition through machine learning techniques offers potent solutions to address the dynamic, complex, and uncertain challenges of commercial site placement. In this study, a machine learning approach was utilized to identify optimal sites, enabling tailored applications to specific business scenarios by adjusting a minimal set of parameters.

2. Relevant Theories

2.1. Overview of the Commercial Site Selection Problem

The approaches which are prevailing in practical apply currently include Analytic Hierarchy Process (AHP), GIS. The main factors taken into calculation and their weights differ depending on distinct types of commercial use.

Taking location selection for market and delivery distribution centers as instance to illustrate the location selection problem.

The factors that effect the location for retail stores were divided as two groups that are social elements and environmental elements by Julie Baker and researchers who set a Matrices model in size of 2*2 according to the importance degree of every factors [5]. In Macroeconomics there are law of retail trade gravitation, central place theory and retail saturation index theory that are widely recognized as cornerstone in classical location selection theory and have been proved effectively in practical application which have been improved by applying genetic algorithm [6-8].

For location selection of delivery distribution centers, customers, supplier distribution, traffic conditions, land conditions, natural conditions and policy conditions is widely considered as the main factors [9]. Probabilistic Model Checking is used in AHP in choosing optimal sites for delivery distribution centers [10].

2.2. Multi-Armed Bandits and Thompson Sampling Algorithm

2.2.1. Multi-Armed Bandits.

Multi-armed bandit (MAB) problem represents a task as a k-armed slot machine. After the arm on the slot machine was pulled a reward is received which is used to estimate the quality of each arm. Higher the reward is present the better arm [11]. The task is to select arms to pull with the goal of maximum the calculated reward after n rounds pulling.

2.2.2. Thompson Sampling Algorithm.

In Bayesian Learning, the uncertainty in the environment is represented by a prior probability distribution which reflect the belief on the environment [12]. In MAB framework the prior distribution update based on the action and reward received in previous step.

The Thompson Sampling Algorithm choose Beta distribution as the prior distribution and current distribution. After choosing an arm a reward presenting as 1 or 0 is received so the distribution update for maximum the cumulative reward.

Thompson Sampling Algorithm is mathematically expressed as.

$$P(\theta) = \frac{P(\frac{D}{\theta})P(\theta)}{P(D)} \tag{1}$$

Where D represents the data observed, P(θ/D) is our posterior, P(D/θ) is the likelihood of observing the data given θ, and P(θ) is the prior belief on the distribution of θ .

The probability density Function is mathematically expressed as

$$f(\theta; \alpha, \beta) = \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \tag{2}$$

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{3}$$

3. System Analysis and Application Research

3.1. Commercial Site Selection Problem Description

Commercial site selection model is used to estimating every candidate address based on the data base and providing them a value for rating. After rating every candidate address is recognized as an arm and goes through the Thompson Sampling process, during which the probability to be chosen in sampling increase if the address perform well in the rating so that after n rounds sampling the optima choice emerge.

Variable kind of data is considered as index for determining the optima site by business site selection model, including population, traffic volume, pedestrian volume, rivals in similar business position, influence in social media, requirements, decoration and maintenance cost, consumption per person in the region, shipping coast and labor cost, whose parameters could be adjusted in the model adopting for applying the model to distinct type of business location accurately and flexibly.

3.2. Mathematical Model

In the usage of Thompson Sampling Algorithm there are m optimal sites that are in the best commercial or transport condition depending on the requirements out of n candidate location. The task of measuring and determining the optimal candidate sites makes the this problem N-P hard in complexity.

In considering of balancing the Exploration and Exploit as the problem could be calculate with the approaches of Multi-armed Bandit, Thompson Sampling Algorithm technique was introduced so that each candidate location was recognized as an arm who could be selected or not in each rounds. With only two possible value could be received in each round, (1 when the chosen location is selected or 0 otherwise) Bernoulli distribution was used to describe the behavior that a reward is received as 1 when the location indexed as successful (which means this location is in good condition estimated by the selection model), or 0 that fails in other condition.

$$\text{if success}: \alpha = \alpha + 1 \tag{4}$$

$$\text{if fail}: \beta = \beta + 1 \tag{5}$$

3.3. Analysis Algorithm

Table 1 is the pseudo code for the selection process in the frame of Thompson Sampling Algorithm :

Table 1. Thompson Sampling Algorithm

for t = 1, 2, 3, … do

Sample the model

for k = 1, .., K do

$\theta_k \sim beta(\alpha_k, \beta_k)$

end for

#selection

$x_t \leftarrow argmax_k \theta_k$

choose x_t,get the reward of r_t

#update the distribution

$(\alpha_{x_t}, \beta_{x_t}) \leftarrow (\alpha_{x_t} + r_t, \beta_{x_t} + 1 - r_t)$

11 end for

According to the strategy process above, the process of commercial site selection problem is :

Taking each candidate site as an arm so there are arm_i,$i \in \{1,2,\dots,m\}$.

Each arm is attached with beta distribution whose initial state is $\beta(1,1)$.

Sampling $\theta_i \sim Beta(\alpha_i, \beta_i)$ and get the value as $(\theta_1, \theta_2, \dots, \theta_k)$.

Choosing the arm owning the largest value,$a_t \leftarrow argmax_k \theta_k$.

Operating to a_t getting the reward r_t.

Updating the distribution $(x_{a_t}, \beta_{a_t}) \rightarrow (\alpha_{a_t} + r_t, \beta_{a_t} + 1 - r_t)$.

Repeating the step (3)(4)(5)(6) until it meets the end.

4. Challenges

4.1. Data Quality and Availability

According to the sensitivity that the Thompson Sampling process perform s when the date fluctuates, a tiny wave in database could be reflected as a huge change in the result, due to which the current data is strictly required otherwise the optimal site could be differ.

The data of pedestrian flow and traffic volume fluctuate greatly depending on the factors of weather, festival and activities around over, which confuse the regular observed in the model.

4.2. Adapting the Algorithm for Diverse Commercial Needs

The business should make it flexible adopting the model to applying in real situation for distinct type of business location. To be specific, in determining the location for restaurants and market the rating model is adjusted to add weight in competitors and pedestrian volume, while the model weights the parameter of transport cost when it is required to choose address for the delivery centers.

5. Conclusion

The macroeconomic model is employed to evaluate potential sites, providing key metrics upon which the Thompson Sampling Algorithm is then applied to select the optimal site. This method is adaptable for various types of commercial locations by fine-tuning parameters associated with influential factors, thus facilitating its application across diverse contexts. Due to its user-friendly operation in practical scenarios, this approach not only offers a novel methodology but also sets a new standard in commercial site selection solutions. With improvements in the accuracy and volume of diverse types of databases, the decision-making model for location selection increasingly enhances the probability of identifying the optimal site for various commercial endeavors. This advancement underscores the growing efficacy of integrating robust data analytics with algorithmic precision to refine and optimize the decision processes in commercial real estate development.

References

[1] J. Li, Statistics in Business Site Selection, China Statistics 163 (2020) 25-26.

[2] Q. Zeng, Research on Business Location Based on Geographic Social Network, Master's thesis, Wuhan University, 2022.

[3] Y. Hu, C. Qin, S. Lv, W. Haojie, Research on the location of electric vehicle charging station considering user satisfaction and range anxiety, Journal of Wuhan University of Technology (Transportation Science and Engineering) (Year) 1-13.

[4] R. Fourie, S.S. Grobbelaar, Healthcare Facility Location Selection: A Bibliometric Analysis and Scoping Review, in: 2022 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), IEEE, 2022, pp. 727-731.

[5] M. Ju, Research on the influencing factors of urban layout of outlets in China, Master's thesis, Shanghai University of Finance and Economics, 2023.

[6] X. Guo, Research on Distribution Center Location-Path Problem Based on Thompson Sampling Algorithm, Master's thesis, Dalian Maritime University, 2023.

[7] J. Guo, Business data analysis of supermarket location based on Arcgis: A case study of Yingze District, Taiyuan City, Shanxi Province, Information and Computer Science (Theoretical Edition) (2020) 34-37.

[8] A. Tarasov, S.J. Delany, B.M. Namee, Dynamic Estimation of Rater Reliability in Subjective Tasks Using Multiarmed Bandits, in: 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing, IEEE, 2012, pp. 979-980.

[9] Z. Teng, J. Zhou, Hub location-allocation problem considering whether passing through hub or not based on Genetic Algorithm, in: 2021 International Conference on E-Commerce and E-Management (ICECEM), IEEE, 2021, pp. 48-52.

[10] X. Sun, C.-C. Wu, L.-R. Chen, Applying analytic hierarchy process to solve distribution center location based on CFLP model, in: 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), IEEE, 2017, pp. 1823-1827.

[11] X. Zhu, Y. Huang, X. Wang, R. Wang, Emotion recognition based on brain-like multimodal hierarchical perception, Multimedia Tools and Applications (2023) 1-19.

[12] J. Baker, M. Levy, D. Grewal, An experimental approach to making retail store environmental decisions, Journal of Retailing 68 (1992) 445-460.