Modeling Commercial Distribution Using Schelling's Model

Omar El Makkaoui

June 2023

Abstract

This study investigates the phenomenon of commercial clustering within urban environments. Despite appearing paradoxical and suboptimal for urban organization, these distributions can be explained through complex dynamics such as those described by Schelling's segregation model. This research models commercial populations using Python, incorporates economic and geographic factors into utility functions, and tests the outcomes against real-world data. The findings aim to inform better urban planning and commercial placement strategies.

1 Introduction

Commercial clustering in certain neighborhoods often appears paradoxical, seeming suboptimal for urban organization. This study explores the factors driving such distributions, focusing on the aggregation dynamics modeled by Schelling's framework. Understanding these phenomena has critical implications for optimizing commercial placements and improving urban planning.

To achieve this, we begin by presenting Schelling's model and its foundational principles, which provide insights into the emergent patterns observed in commercial distributions. A simplified approach is initially adopted to model aggregation dynamics, enabling a clear observation of clustering behavior. Building upon this foundation, we explore the practical application of the model to commercial distribution, considering both deterministic and stochastic approaches. The selection of an appropriate utility function is crucial in capturing the economic and geographic realities of urban environments, and various optimization techniques such as gradient descent are employed to refine the model's accuracy. Additionally, implementation efforts are directed towards applying these methodologies to real-world scenarios to assess their practical relevance.

The study also delves into the stochastic approaches and limitations of the model. Techniques such as the Metropolis-Hastings algorithm are utilized to discretize the problem, while numerical simulations using the Monte Carlo method provide further insights into the system's behavior. A critical evaluation of these approaches allows for a deeper understanding of the strengths and weaknesses of the model, informing potential improvements and future research directions.

The thematic scope of this research spans:

- **Mathematics:** Analysis and applied mathematics.
- **Computer Science:** Practical applications of computational methods.

1.1 Problem Statement

Research Questions:

- How can the location of a new business within the city be optimized?
- What are the different phenomena that emerge during the distribution of businesses, and can they be quantified?

2 Literature Review

The placement of commercial establishments is a crucial economic decision for businesses aiming to maximize profitability. Factors such as population income and proximity to competing businesses play key roles in this optimization process.

Schelling's model [1] illustrates the consequences of uncoordinated yet interdependent behaviors, where individual decisions can lead to collective phenomena such as clustering. Forsé and Parodi [2] further emphasize the interplay between individual preferences and collective outcomes. Additionally, Farahani and Hekmatfar [3] provide insights into facility location optimization.

Keywords

Commercial clustering, Schelling's model, urban planning, business location optimization, stochastic modeling, utility functions, gradient descent, Monte Carlo method, Metropolis-Hastings algorithm.

3 Methodology

3.1 Schelling's Model

The spatial segregation model proposed in the 1970s by Thomas C. Schelling is a simplified model that highlights the aggregation resulting from population dynamics. This model demonstrates that despite neutral individual choices, collective effects lead to population segregation.

In the example I chose The model considers two different populations in equal proportions. Each agent in the model possesses a utility function that depends solely on its closest neighbors. If an agent finds that its utility function is higher elsewhere, it relocates to that location. This mechanism captures the underlying tendencies driving commercial clustering within urban environments. By applying this framework, the study aims to simulate commercial distributions and analyze their emergent properties.

3.2 Application of Schelling's Model to Commercial Distribution

This section focuses on the application of Schelling's model to the distribution of businesses, examining both deterministic and stochastic approaches. A key aspect of this application involves selecting an appropriate utility function, which is a personal initiative in this study. Furthermore, minimization methods, particularly analytical approaches such as gradient descent, are explored. Finally, the implementation of gradient descent techniques for commercial placement is discussed to enhance the model's practical applicability.

3.3 Factors Influencing Store Location

In order to determine the appropriate utility function for commercial distribution, it is essential to analyze the factors influencing this dynamic.

These factors are divided into two categories. For the purpose of our study, we will focus only on the gravitational attraction of the store.

Key Influencing Factors:

3.3.1 Gravitational Attraction

Gravitational attraction plays a crucial role in determining store placement. It depends on:

- The geographical proximity of the store to its clientele.

- The size of the store.

3.3.2 Attitudinal Attraction

Attitudinal attraction influences customer choice based on:

- Price-quality ratio.
- Service quality.
- Staff friendliness.

3.4 Deterministic Approach: Analogy with Celestial Mechanics

By analogy with the universal law of gravitation, a possible modeling approach would be to assume that:

- The population represents the equivalent of mass in physics;
- Each store exerts a "gravitational" attraction on individuals located nearby;
- Each individual is in a gravitational field where competing forces of different intensities and directions interact.

The deterministic approach draws an analogy with celestial mechanics, where:

Population	Mass
Store	Star
Utility function	Gravitational attraction

Interaction between stores:

$$T_{ij} = K \frac{W_i^{(1)} W_j^{(2)}}{C_{ij}^n} \tag{1}$$

3.5 Stochastic Approach: Huff's Model

The stochastic approach is more suited to reality. We will rely on the Hull model introduced in the early 1960s.

According to Hull, a consumer is not irrevocably tied to a store but is likely to hesitate between several shopping locations. Therefore, all stores have a chance of being visited, contrasting with the deterministic approach that prevailed at the time.

$$P_{ij} = \frac{U_{ij}}{\sum_{k \in N_i} U_{ik}} \tag{2}$$

where the utility of the store is:

$$U_{ij} = S_j^{\alpha} D_{ij}^{\beta} \tag{3}$$

The parameters α and β reflect the importance given to store size and distance in the consumer's decision to visit a particular store. Since utility decreases with distance, the parameter is negative. The size S_j of store j can be replaced with any other measure of the store's attractiveness.

3.6 Choice of the Utility Function

To determine the appropriate utility function for commercial distribution, it is necessary to analyze the influencing factors. A simplified approach is considered to facilitate the evaluation.

General Formula:

$$f(X) = \sum_{i=1}^{m} w_i d(X, P_i) \tag{4}$$

Squared Euclidean Distance:

$$f(X) = \sum_{i=1}^{m} w_i \left[(x - a_i)^2 + (y - b_i)^2\right] \tag{5}$$

Minimization of the Utility Function:

$$\left(\frac{\partial f(x,y)}{\partial x}, \frac{\partial f(x,y)}{\partial y}\right)\Bigg|_{(x=x^*, y=y^*)} = (0,0) \tag{6}$$

With:

$$x^* = \frac{\sum_{i=1}^{m} w_i a_i}{\sum_{i=1}^{m} w_i} \tag{7}$$

$$y^* = \frac{\sum_{i=1}^{m} w_i b_i}{\sum_{i=1}^{m} w_i} \tag{8}$$

where:

- w_i: weight assigned to each store, ranging between 0 and 1.
- P_i: position of other stores.
- X: position of the studied store.

3.7 Simplified Utility Function (All or Nothing Form)

Simplified Form: ALL OR NOTHING

$$\Gamma(M, x) = \iint_K \mathbf{1}[d(y, x) \leq d(y, M)]\rho(y)\, dy_1 dy_2 \tag{9}$$

where:

- M: a finite set representing the positions of existing stores.
- $\rho(y)$: population density.
- K: a compact subset of $\mathbb{R}^2$.

3.8 General Form of the Utility Function

General Form:

$$\Gamma(M, x) = \iint_K f(M, x, y)\rho(y)\, dy_1 dy_2 \tag{10}$$

where:

$$f(\mathcal{M}, x, y) = \frac{h(d(y, x))}{\sum_{\tilde{M}:=\mathcal{M}\cup\{x\}} h(d(y, z))} \tag{11}$$

$$h_\lambda : d \to e^{-\lambda d} \tag{12}$$

and:

$$\lambda \gg d \tag{13}$$

Choosing $h_\lambda : d \to e^{-\lambda d}$ implies that if we select λ large relative to the characteristic distances between stores, Γ will approximate an indicator function. This is because, in the denominator of f, a single term—corresponding to the closest store (real or hypothetical)—will dominate all others. As a result, f will be very close to either 0 or 1.

3.9 Gradient Descent Application to Our Case

Gradient descent is an approximate optimization algorithm used to find the minimum of any differentiable function. It is an iterative algorithm that starts from an arbitrarily chosen point and moves in the direction of the steepest descent of the function to be minimized, which corresponds to the direction of the negative gradient.

Gradient Calculation:

$$\nabla\Gamma(M, x) = \iint_K \nabla \left(\frac{h(d(x, y))}{\sum_{z \in M \cup \{x\}} h(d(z, y))} \right) \rho(y)\, dy_1 dy_2 \tag{14}$$

$$\nabla\Gamma(M, x) = \iint_K h'(d(x, y)) \frac{\sum_M h(d(y, z))}{\left(\sum h(d(z, y)) + h(d(x, y))\right)^2} \nabla d(y, x) \rho(y)\, dy_1 dy_2 \tag{15}$$

$$\nabla\Gamma(M, x) = \int_0^{100} \int_0^{100} h'(d(x, y)) \frac{\sum_M h(d(y, z))}{\left(\sum h(d(z, y)) + h(d(x, y))\right)^2} \nabla d(y, x) \rho(y)\, dy_1 dy_2 \tag{16}$$

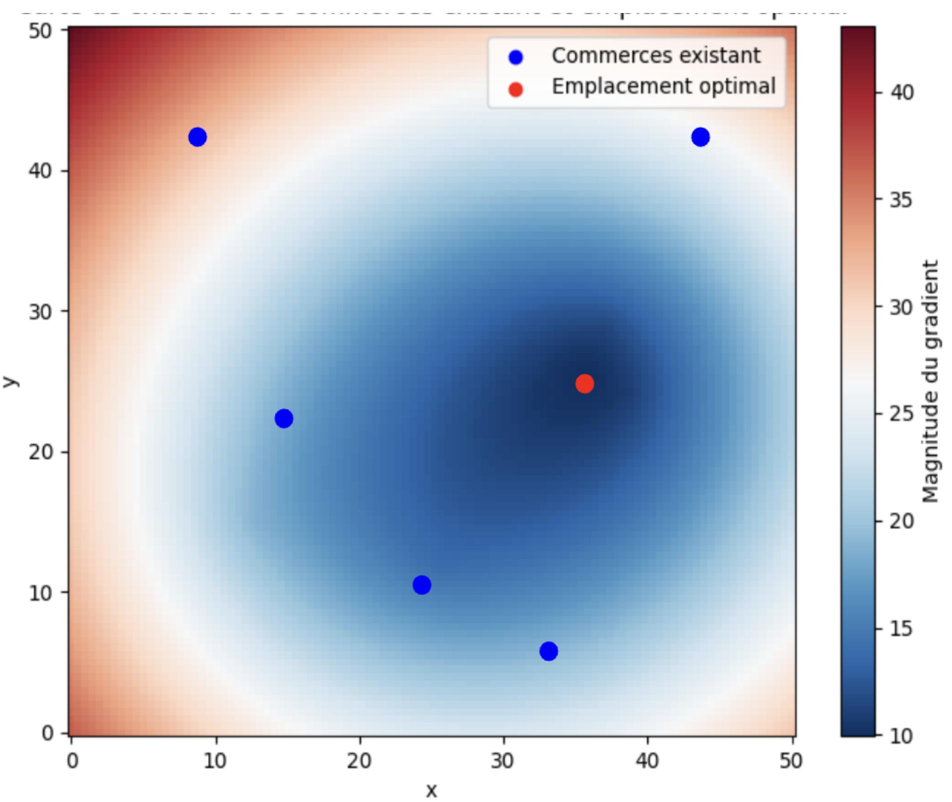

Figure 1: Gradient descent for one store in the presence of five existing stores.

4 Stochastic Approach and Model Limitations

To address the inherent uncertainty and randomness present in commercial distribution, a stochastic approach is considered. This section explores numerical methods that provide a deeper understanding of the probabilistic nature of business clustering and the associated limitations.

One of the key techniques employed for the discretization of the problem is the **Metropolis-Hastings algorithm**, which is widely used in Markov Chain Monte Carlo (MCMC) simulations. This algorithm offers an effective way to explore the solution space by iteratively sampling potential commercial placements, ensuring convergence towards an optimal distribution over time.

Furthermore, a numerical approach based on the **Monte Carlo method** is applied to evaluate the distribution of commercial establishments. The Monte Carlo simulation, implemented as a personal initiative, involves generating a large number of random samples to approximate potential configurations of commercial locations. This probabilistic technique provides valuable insights into the robustness and reliability of the model under various scenarios.

Finally, the **perspective on modeling limitations** is discussed, emphasizing the constraints and challenges encountered in the implementation process. While stochastic approaches offer flexibility and adaptability, they also introduce computational complexity and require careful calibration to ensure meaningful and actionable results in practical applications.

4.1 Definition of the Gibbs Measure

We define the Gibbs measure μ_T which assigns a probability to each site in E and concentrates on the minima of V as T tends to zero. The following result is a simplified version of Laplace's method.

$$\forall x \in E, \quad \mu_T(x) = \frac{1}{Z_T} \exp\left(\frac{-V(x)}{T}\right)$$

with

$$Z_T = \sum_{y \in E} \exp\left(\frac{-V(y)}{T}\right)$$

where:

- V: Utility function to be minimized or maximized.
- E: Set of positions accessible to the business.

4.1.1 Analogy with Statistical Thermodynamics

According to Boltzmann's law, the probability for a given particle to occupy the quantum state of energy E_i is:

$$p(E_i) = \frac{1}{Z} \exp\left(\frac{-E_i}{k_b T}\right) \quad \text{with } Z = \sum_i \exp\left(\frac{-E_i}{k_b T}\right)$$

4.2 Determination of Minima

Laplace's Lemma:

$$\forall x \in E, \quad \lim_{T \to 0} \mu_T(x) = \frac{1}{\text{Card}(M)} \mathbf{1}(x \in M)$$

For values of T close to zero, the Gibbs measure μ_T concentrates on points where V reaches its minimum. Consequently, by simulating realizations of the measure μ_T for T near zero, we will obtain, with high probability, an approximation of the set M where V reaches its minimum.

5 Implementation of Gibbs Measure Simulation

At first glance, simulating the Gibbs measure:

$$\mu_T(x) = \frac{1}{Z_T} \exp\left(-\frac{V(x)}{T}\right)$$

assumes computing the distribution μ_T by evaluating:

$$Z_T = \sum_{y \in E} \exp\left(-\frac{V(y)}{T}\right)$$

In practice, this is impossible to implement because it would require evaluating V for all values in a large set E.

5.1 The Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm overcomes this difficulty by simulating the Gibbs measure using a Markov chain. The algorithm allows for the simulation of a random variable under any probability measure.

Algorithm Idea:

- Start from a randomly chosen point x.
- Choose another random point y.
- The algorithm then decides whether to move to y with probability p.
- The probability p is determined by a function that maximizes p when $V(x) < V(y)$.

This feature allows the algorithm to escape local maxima and reach the global maximum.

6 Markov Chains

6.1 Markov Property

A sequence of random variables $\{X_n\}_{n \geq 0}$ taking values in a set E satisfies the Markov property:

$$\mathbb{P}(X_{n+1} = y \mid X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = \mathbb{P}(X_{n+1} = y \mid X_n = x_n)$$

6.2 Transition Matrix

The transition matrix from state x to y is defined as:

$$\forall x, y \in E, \quad P(x, y) = \mathbb{P}(X_{n+1} = y \mid X_n = x)$$

7 Steps for Simulation Implementation

7.1 Transition Matrix Symmetry

We consider a symmetric transition matrix such that:

$$Q(x, y) = Q(y, x)$$

A monotonous function satisfying the following condition is used:

$$h:]0,\infty[\to [0,1] \quad \text{such that } h(u) = uh(1/u)$$

For example:

$$h(u) = \frac{u}{1+u}$$

7.2 Simulation Steps

Initialization:

- We initialize X_0 randomly.

Inheritance:

- Choose y according to the law $Q(X_n, y)$.
- Select U_{n+1} uniformly from $[0,1]$, independently of the previous values.
- If $U_{n+1} > h\left(e^{\frac{V(X_n)-V(y)}{T_n}} \frac{Q(y,X_n)}{Q(X_n,y)}\right)$, then set $X_{n+1} = y$.
- Otherwise, set $X_{n+1} = X_n$.

Theorem 2: Given a potential function V and a transition matrix Q over a finite state space E, there exists a constant $C(V,Q)$ such that the simulated annealing algorithm with the temperature sequence

$$T_n = \frac{C(V,Q)}{\log(n)}$$

selects the set M of minima of V, meaning that:

$$\lim_{n\to\infty} \mathbb{P}(X_n \in M) = 1$$

7.3 Monte Carlo Method for the Approximate Calculation of the Utility Function

Utility Function:

$$\Gamma(M,x) = \iint_K f(M,x,y)\rho(y)\, dy_1 dy_2$$

One-Dimensional Case:

Uniform probability density over the interval [a,b] **of a random variable** X:

$$p(x) = \frac{1}{b-a}$$

Evaluation of the integral using the Monte Carlo method by taking N samples of X:

$$S_N = (b-a)\frac{1}{N}\sum_{i=0}^{N-1} f(x_i) \approx \int_a^b f(x)\,dx$$

8 Results and Discussion

The initial simulations demonstrate that even neutral individual decisions can lead to significant clustering due to collective dynamics. The Schelling model effectively captures the emergence of commercial clusters despite the absence of centralized coordination, highlighting the importance of individual utility functions in shaping urban landscapes.

8.1 Key Findings

- **Geographic Influence:** The simulations confirm that geographic proximity plays a crucial role in clustering intensity. Areas with higher population density tend to attract more commercial establishments, validating the gravitational attraction hypothesis.
- **Economic Parameters:** Factors such as purchasing power and competitive pricing influence the stabilization of distribution patterns. The introduction of economic constraints in the utility function refines the model's predictive power.
- **Gradient Descent Optimization:** The gradient descent algorithm efficiently locates optimal store placements, minimizing the utility function and offering insights into potential commercial hotspots.
- **Gibbs Measure Interpretation:** The simulations align with the theoretical expectations from the Gibbs measure approach. As temperature decreases, the model concentrates on areas with minimal utility values, providing a probabilistic framework for optimal commercial placement.
- **Monte Carlo Approximations:** The Monte Carlo method effectively approximates complex integrals related to utility functions, validating its use in large-scale simulations of commercial distribution.
- **Metropolis-Hastings Algorithm Efficiency:** This stochastic method successfully bypasses the infeasibility of exhaustive computations by generating representative samples from the Gibbs distribution. It enables the exploration of commercial placement strategies without requiring complete knowledge of the entire state space.

8.2 Comparison with Real-World Data

Comparing the model's predictions with real-world commercial distributions reveals a strong correlation between predicted optimal locations and actual commercial hotspots. However, some discrepancies arise due to:

- The oversimplification of social and psychological factors influencing customer behavior.
- The assumption of uniform population density, which does not always hold in practical scenarios.
- The static nature of the model, which does not account for temporal dynamics such as seasonal demand fluctuations.

8.3 Limitations and Future Work

While the model provides valuable insights, several limitations must be addressed:

- **Dynamic Considerations:** Future work should incorporate temporal elements to capture evolving commercial dynamics over time.

- **Behavioral Complexity:** Including more sophisticated consumer behavior models could enhance accuracy.
- **Urban Planning Integration:** Collaborating with urban planners to incorporate additional zoning and infrastructure constraints.
- **Refinement of Parameters:** Further calibration using high-resolution demographic and economic data to enhance predictive accuracy.

Overall, this research demonstrates the potential of Schelling's model in analyzing and optimizing commercial distribution while also highlighting areas for improvement to better align with real-world complexities.

9 Conclusion

This research underscores the utility of Schelling's model in understanding commercial distributions. The findings provide valuable insights for urban planning and business decision-making. Future work will focus on incorporating dynamic market trends and refining utility functions for greater accuracy.

References

References

[1] Schelling, T. (1971). Dynamic Models of Segregation. *Journal of Mathematical Sociology*, 1(2), 143–186.

[2] Forsé, M., & Parodi, M. (2014). Spatial Segregation According to Schelling: Perversity Lies Elsewhere. `https://hal-sciencespo.archives-ouvertes.fr/hal-00973079/document`

[3] Farahani, R., & Hekmatfar, M. (2010). *Facility Location: Concepts, Models, Algorithms and Case Studies.*

[4] Baray, J. (2015). Multiple Commercial Location: Application of Signal Processing and p-Median Model to Developing Organic Product Networks. `https://hal.science/tel-01121627v1/document`

[5] Dentan, J., Le Roux, S., & Draznieks, G. (2021). Dynamics of a Commercial Population on the Saclay Plateau. `https://www.researchgate.net/publication/351102583_Dynamics_of_a_Commercial_Population_on_the_Saclay_Plateau`