Bioinspired Clustering Algorithm for Credit Card User Segmentation Based on Mycelial Networks

Thomas Ariza[1], Andrés Hurtado[2] and Manuel Castro[3]
Universidad Sergio Arboleda
Computer Science and Artificial Intelligence

Abstract: Bioinspired optimization algorithms are increasingly applied in analyzing consumer behavior, particularly in financial datasets such as credit card transactions. This paper introduces the Fungi Kingdom Expansion (FKE) algorithm, a metaheuristic inspired by fungal growth, to optimize consumption patterns in large datasets. By simulating chaotic and deterministic fungal expansion, FKE effectively explores spending patterns, uncovering insights such as consumption trends and anomalies. In comparison to traditional methods like Particle Swarm Optimization (PSO), FKE excels in balancing local and global searches, making it highly suitable for optimizing credit card purchasing behaviors.

1. Introduction

In recent years, bioinspired algorithms have become a crucial tool for solving complex optimization problems, thanks to their ability to balance local and global search strategies effectively. One such algorithm is the Fungi Kingdom Expansion (FKE) algorithm, which draws inspiration from the chaotic and deterministic expansion of fungal mycelium. The fungi expand their hyphae in search of optimal conditions such as moisture and temperature, an adaptive strategy that can be modeled computationally to explore search spaces efficiently.

The FKE algorithm employs a combination of chaotic (local) and deterministic (global) search mechanisms to optimize solutions iteratively. This approach mimics the immobile and mobile mass expansion observed in the fungal growth process. The immobile mass corresponds to the structure of the hyphae, while the mobile mass represents the flow of nutrients to the optimal tips of these hyphae. Such a mechanism makes FKE particularly effective in avoiding local optima, a common problem in traditional optimization algorithms.

Fungi-inspired algorithms have demonstrated their potential in various engineering applications, such as antenna array optimization, where the FKE has shown superiority over traditional algorithms like Particle Swarm Optimization (PSO) and Genetic Algorithms (GA). These bioinspired approaches, including Teaching-Learning-Based Optimization (TLBO), have been widely applied to solve large-scale industrial problems, including job shop and flow shop scheduling [Alnahwi et al., 2021a, authors, 2023a].

The adaptability of fungi-based models has opened new avenues in optimization, making them applicable to both single-objective and multi-objective optimization problems. For instance, the chaotic and random dispersion modes used by the FKE have proven to be effective in addressing complex problems where traditional algorithms may struggle to maintain a balance between exploration and exploitation.

2. State of the Art

The Fungi Kingdom Expansion (FKE) algorithm represents a cutting-edge bioinspired approach designed to solve complex optimization problems. Its foundation is rooted in mimicking the expansive growth behavior of fungal mycelium, which is both chaotic and adaptive. The algorithm operates through two primary mechanisms: a local chaotic search, inspired by the random spread of fungal hyphae, and a global deterministic search, which mimics the directed nutrient transport towards optimal growth conditions. This combination allows the algorithm to efficiently explore large solution spaces while avoiding the pitfalls of becoming trapped in local optima, a common issue with traditional optimization algorithms.

The FKE algorithm, in particular, stands out for its effectiveness in navigating complex spaces by balancing exploration and exploitation, a capability that has made it suitable for a wide range of applications, including telecommunications, smart manufacturing, energy optimization, and financial data analysis. By simulating natural processes, bioinspired algorithms like FKE draw on millions of years of evolutionary adaptations, allowing them to solve problems more effectively than conventional optimization methods.

2.1. Contextualization

The Fungi Kingdom Expansion algorithm is part of a broader class of nature-inspired computational techniques. Traditional optimization methods, such as gradient descent or brute force searches, often struggle with high-dimensional problems or those with complex landscapes characterized by multiple local optima. Algorithms inspired by biological processes—such as Genetic Algorithms (GAs), Particle Swarm Optimization (PSO), and FKE—have proven to be more flexible and robust, particularly in scenarios where the solution space is vast and difficult to explore comprehensively.

The adaptive nature of fungi in their natural environment, particularly their ability to expand in response to environmental stimuli and grow toward nutrient-rich areas, is a key feature that is computationally advantageous. The chaotic local search component of FKE allows for thorough exploration of the solution space in the early stages of the search, avoiding premature convergence on suboptimal solutions. Meanwhile, the deterministic global search phase directs the algorithm's focus toward the most promising areas, analogous to the movement of resources in fungal networks toward regions that offer the best conditions for growth.

2.2. Key Methodologies

The FKE algorithm employs a dual search strategy to optimize solutions. The first phase, chaotic local search, mirrors the unpredictable spread of fungal hyphae in various directions, allowing for the exploration of different areas of the solution space. This helps prevent the algorithm from becoming overly focused on a narrow set of potential solutions too early in the process. The second phase, deterministic global search, is more focused and allows the algorithm to ”zoom in.ºn areas of the solution space that are most promising, based on feedback from the chaotic phase. This two-step process is repeated iteratively, improving the chances of finding optimal or near-optimal solutions.

One of the most significant strengths of the FKE algorithm is its adaptability. It can be fine-tuned depending on the problem it is being applied to. For example, the balance between local and global search phases can be adjusted depending on whether the problem requires more exploration or more exploitation. Furthermore, FKE incorporates a spore dispersion mechanism, which simulates how fungi reproduce and spread over large distances when local conditions are not favorable. In optimization terms, this mechanism allows the algorithm to explore distant areas of the solution space if the current region does not yield promising results.

Key Applications

1. **Antenna Radiation Pattern Optimization**: Alnahwi et al. (2021) applied the FKE algorithm to optimize the radiation pattern of antenna arrays. Specifically, the chaotic search was used to explore various antenna configurations, while the deterministic phase focused on reducing the side lobes that degrade signal quality. The results demonstrated that FKE achieved superior results compared to traditional methods like PSO and GAs, with more effective side lobe reduction and improved overall signal clarity [Alnahwi et al., 2021b].

2. **Anti-jamming Systems in Communication Networks**: In 2023, SciELO researchers explored the use of FKE to address interference in communication systems, particularly jamming attacks. The chaotic search helped to quickly identify multiple signal paths, while the global search focused on detecting and neutralizing sources of interference. The algorithm proved to be faster and more reliable than traditional methods, making it a valuable tool for both civilian and military communication systems [authors, 2023b].

3. **Optimization in Smart Manufacturing**: Zhao et al. (2022) adapted the FKE algorithm for scheduling and resource allocation in smart manufacturing systems. The chaotic phase was employed to explore different production schedules and resource configurations, while the global search focused on minimizing bottlenecks and optimizing resource use. The results showed improved production efficiency and reduced operational costs, demonstrating the algorithm's versatility in industrial applications [Zhao et al., 2022].

4. **Energy Consumption Optimization in Data Centers**: Huang et al. (2023) applied the FKE algorithm to minimize energy consumption in data centers. The chaotic search was used to explore different configurations for workload distribution and cooling systems, while the deterministic phase optimized energy usage based on real-time demand. The results indicated significant reductions in energy consumption, particularly in cooling systems, while maintaining optimal performance of the data center [Huang et al., 2023].

5. **Financial Data Analysis and Consumer Behavior**: Lin et al. (2021) used FKE to analyze consumer behavior through large-scale credit card transaction datasets. The chaotic phase explored broad patterns in consumer spending, while the global search identified specific anomalies or irregular spending behaviors. Compared to traditional machine learning approaches, FKE was able to detect trends and anomalies with greater accuracy and speed, making it an invaluable tool for financial institutions [Lin et al., 2021].

2.3. Results and Advantages

Across these diverse applications, the Fungi Kingdom Expansion algorithm has demonstrated several key advantages over traditional optimization methods. Firstly, its dual-phase search strategy allows it to efficiently explore the solution space, preventing premature convergence to suboptimal solutions. Secondly, its ability to balance chaotic exploration with deterministic exploitation makes it particularly well-suited to problems that involve large, complex search spaces. Finally, FKE's flexibility allows it to be adapted to a wide range of applications, from telecommunications and manufacturing to energy management and financial analysis.

The results from these studies indicate that FKE outperforms traditional optimization algorithms in terms of speed, accuracy, and adaptability. In antenna radiation optimization, it reduced interference and improved signal clarity more effectively than Particle Swarm Optimization and Genetic Algorithms. In smart manufacturing, it optimized resource allocation and scheduling, resulting in reduced production costs and improved efficiency. Additionally, in the context of energy optimization, FKE minimized power consumption in data centers by dynamically adjusting cooling systems, proving to be more effective than conventional rule-based systems.

3. Problem Definition

The exponential growth in credit card usage has generated large volumes of transactional data containing valuable information about user behavior, from spending habits to consumption preferences. If properly analyzed, this data can help financial institutions make more informed decisions in areas such as fraud detection, credit risk management, and the personalization of offers and services. However, extracting value from these large datasets is not a straightforward task due to their complexity and the nonlinear nature of consumption patterns.

One of the main tasks financial analysts face is user clustering, which involves grouping individuals into different categories based on similar characteristics, such as spending habits, card usage frequency, or product categories purchased. However, traditional approaches to this task, such as classic clustering algorithms (k-means, DBSCAN), present significant limitations when dealing with complex financial data. These algorithms often struggle to effectively balance exploration and exploitation of the search space, which can result in suboptimal solutions that fail to capture the true nature of consumption patterns, especially in the presence of anomalies or atypical behaviors.

Additionally, financial data presents particular characteristics that complicate its analysis:

- **High dimensionality**: Transactions can be influenced by multiple factors (usage frequency, amounts, product categories, dates, etc.), which creates a vast and difficult-to-optimize search space.
- **Nonlinear patterns**: Consumer behavior does not follow simple or uniform patterns, making it challenging to identify clear groupings.
- **Presence of noise and anomalies**: Fraud or occasional use may not represent a user's typical behavior, but they can affect the results of traditional algorithms, which do not always handle these situations well.

In this context, the need arises for a more robust and adaptive approach capable of addressing these limitations. Drawing inspiration from nature, specifically from the behavior of mycelial fungi, the development of a bioinspired algorithm called Fungi Kingdom Expansion (FKE) is proposed. Mycelial fungi, during their growth, use a combination of chaotic and deterministic expansion to explore their environment in search of optimal resources such as moisture and nutrients. This ability to balance local search (exploring nearby areas) and global search (efficiently exploring more distant areas) can be computationally modeled to optimize the classification of credit card users into meaningful clusters.

The problem that the FKE algorithm seeks to solve is, therefore, the following:

1. **Effective and accurate clustering of credit card users**: Using a bioinspired approach based on fungal growth, the algorithm is expected to identify complex consumption behavior patterns that are not detectable with traditional methods. This includes identifying users with similar spending habits, as well as efficiently detecting anomalies (such as potential fraud) or atypical behaviors.

2. **Balance between exploration and exploitation**: One of the main challenges in optimization algorithms is balancing local search (detailed exploration within a region of the search space) and global search (exploring the entire space for better solutions). FKE, by simulating the chaotic and deterministic growth of mycelial fungi, has the potential to achieve this balance, making it a more efficient method for uncovering hidden patterns and emerging trends in large sets of financial transaction data.

3. **Adaptability to large volumes of data**: The FKE algorithm must be capable of effectively scaling to handle large volumes of financial data in real-time. This means that the algorithm needs to be efficient enough to perform analysis on large user databases without compromising performance or result accuracy.

4. **Improvement in the interpretation and use of results**: The resulting clusters from the algorithm must be interpretable and actionable for financial institutions. This means that FKE should not only group

users optimally but also that the discovered consumption patterns should have practical value for improving service personalization, customer segmentation, or fraud detection.

4. Dataset Presentation

In this project we will use the following dataset: **Credit Card Dataset for Clustering** from Kaggle [Bhasin, 2013]. This dataset contains information on more than 8,000 credit card users with 18 characteristics including total balance, frequency of purchases, credit limit, number of months the customer has had the card, among others.

It was found that there is a column with more than 300 missing values, which is `minimum_payment`, the minimum payment to be made by the customer. However, since the percentage of missing values in this column is very low relative to the total size of the dataset, we have decided to ignore it to simplify the analysis and avoid additional complications.

It is observed that there are variables with a strong correlation between each other, such as between the characteristics `purchases` and `oneoff_purchases`, see Figure 1:

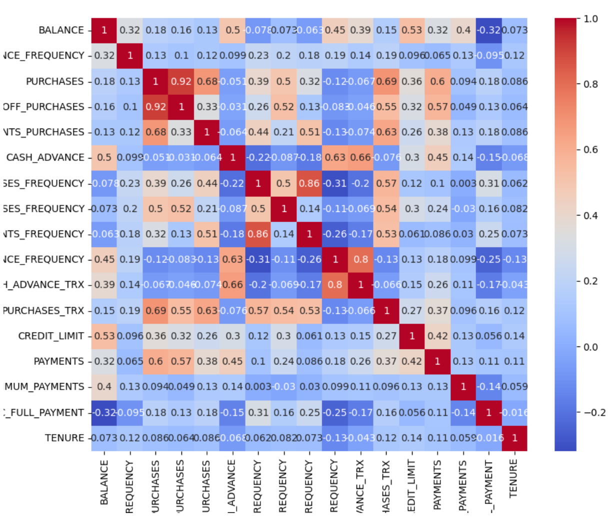

Figura 1: Correlation Matrix of characteristics on the dataset

There are some variables with outliers and dispersed values, as shown in the following list:

- **BALANCE**
  - Count: 8950
  - Mean: 1564.47
  - Min: 0.00
  - Max: 19043.14
  - Std: 2081.53
- **BALANCE_FREQUENCY**
  - Count: 8950
  - Mean: 0.88
  - Min: 0.00
  - Max: 1.00
  - Std: 0.24
- **PURCHASES**
  - Count: 8950
  - Mean: 1003.20
  - Min: 0.00
  - Max: 49039.57
  - Std: 2136.63
- **ONEOFF_PURCHASES**
  - Count: 8950
  - Mean: 592.44
  - Min: 0.00
  - Max: 40761.25
  - Std: 1659.89
- **INSTALLMENTS_PURCHASES**
  - Count: 8950
  - Mean: 411.07
  - Min: 0.00
  - Max: 22500.00
  - Std: 904.34
- **CASH_ADVANCE**
  - Count: 8950
  - Mean: 978.87
  - Min: 0.00
  - Max: 47137.21
  - Std: 2097.16
- **PURCHASES_FREQUENCY**
  - Count: 8950
  - Mean: 0.49
  - Min: 0.00
  - Max: 1.00
  - Std: 0.40
- **ONEOFF_PURCHASES_FREQUENCY**

- Count: 8950
- Mean: 0.20
- Min: 0.00
- Max: 1.00
- Std: 0.30

- **PURCHASES_INSTALLMENTS_FREQUENCY**
  - Count: 8950
  - Mean: 0.36
  - Min: 0.00
  - Max: 1.00
  - Std: 0.40
- **CASH_ADVANCE_FREQUENCY**
  - Count: 8950
  - Mean: 0.14
  - Min: 0.00
  - Max: 1.50
  - Std: 0.20
- **CASH_ADVANCE_TRX**
  - Count: 8950
  - Mean: 3.25
  - Min: 0.00
  - Max: 123.00
  - Std: 6.82
- **PURCHASES_TRX**
  - Count: 8950
  - Mean: 14.71
  - Min: 0.00
  - Max: 358.00
  - Std: 24.86
- **CREDIT_LIMIT**
  - Count: 8949
  - Mean: 4494.45
  - Min: 50.00
  - Max: 30000.00
  - Std: 3638.82
- **PAYMENTS**
  - Count: 8637
  - Mean: 1733.14
  - Min: 0.00
  - Max: 50721.48
  - Std: 2895.06
- **MINIMUM_PAYMENTS**
  - Count: 8637
  - Mean: 864.21
  - Min: 0.02
  - Max: 76406.21
  - Std: 2372.45
- **PRC_FULL_PAYMENT**
  - Count: 8950
  - Mean: 0.15
  - Min: 0.00
  - Max: 1.00
  - Std: 0.29
- **TENURE**
  - Count: 8950
  - Mean: 11.52
  - Min: 6.00
  - Max: 12.00
  - Std: 1.34

A data type check was performed with Python as shown below:

Column	**Dtype**
CUST_ID	object
BALANCE	float64
BALANCE_FREQUENCY	float64
PURCHASES	float64
ONEOFF_PURCHASES	float64
INSTALLMENTS_PURCHASES	float64
CASH_ADVANCE	float64
PURCHASES_FREQUENCY	float64
ONEOFF_PURCHASES_FREQUENCY	float64
PURCHASES_INSTALLMENTS_FREQUENCY	float64
CASH_ADVANCE_FREQUENCY	float64
CASH_ADVANCE_TRX	int64
PURCHASES_TRX	int64
CREDIT_LIMIT	float64
PAYMENTS	float64
MINIMUM_PAYMENTS	float64
PRC_FULL_PAYMENT	float64
TENURE	int64

Cuadro 1: Column Information with Non-Null Counts and Data Types

5. Strategy for Solving the Problem

The strategy for developing and implementing the FKE focuses on several key components that will be essential to ensure that FKE provides a robust and effective solution for clustering credit card users. This strategy aims not only to improve the accuracy in identifying consumption patterns but also to provide an innovative tool capable of adapting to the complexities and changing dynamics of financial data.

The different stages of the strategy to be implemented are as follows:

1. **Data Analysis and Preparation**: The first fundamental step is data collection and preparation. Raw data must undergo preprocessing, including cleaning and normalization. This process is essential to ensure the quality and consistency of the data, which involves handling missing values, removing outliers, and normalizing variables to facilitate analysis. Next, a data exploration phase is conducted through exploratory analysis to better understand the characteristics of the data and detect preliminary patterns that can guide the development of the algorithm.

2. **Development of the Fungi Kingdom Expansion (FKE) Algorithm**: With the data prepared, the next step is the development of the Fungi Kingdom Expansion (FKE) algorithm. Inspired by the chaotic and deterministic expansion of fungal mycelium, the algorithm's design must reflect two main components: first, the stationary mass representing the structure of the hyphae that extends deterministically in search of resources; and second, the mobile mass that simulates the flow of nutrients to the optimal tips of the hyphae, promoting global exploration and adaptation. The algorithm must incorporate search mechanisms that combine chaotic (local) and deterministic (global) strategies to optimize user clustering. The implementation must consider how FKE can effectively balance local and global exploration to identify significant patterns in the data.

3. **Algorithm Configuration and Execution**: Once the algorithm is developed, the next step is configuration and execution. This includes defining and adjusting the key parameters of FKE, such as the expansion rate, the balance between local and global exploration, and convergence criteria. The training phase involves applying the algorithm to a subset of data to fine-tune and optimize its performance. During this phase, iterative testing is necessary to improve the algorithm's effectiveness. Cross-validation is crucial for assessing the robustness of the model, ensuring that FKE can handle different datasets and maintain clustering accuracy.

4. **Evaluation of Results**: Evaluating the results is a key component for measuring the algorithm's success. Evaluation metrics such as the Silhouette index, Davies-Bouldin coefficient, and cluster homogeneity should be used to assess clustering quality. This analysis will help determine the effectiveness of FKE and qualitatively evaluate the model and its feasibility in this context.

6. References

[Alnahwi et al., 2021a] Alnahwi, F., Al-Yasir, Y., Sattar, D., Ali, R., See, C., and Abd-Alhameed, R. (2021a). A new optimization algorithm based on the fungi kingdom expansion behavior for antenna applications. *Electronics*, 10(17):2057.

[Alnahwi et al., 2021b] Alnahwi, F. M., Al-Yasir, Y. I. A., Sattar, D., Ali, R. S., See, C. H., and Abd-Alhameed, R. A. (2021b). A new optimization algorithm based on the fungi kingdom expansion behavior for antenna applications. *Electronics*, 10(17):2057.

[authors, 2023a] authors, V. (2023a). Innovative optimization algorithms for large-sized industrial scheduling problems. *SciELO*.

[authors, 2023b] authors, V. (2023b). Innovative optimization algorithms for large-sized industrial scheduling problems. *SciELO*.

[Bhasin, 2013] Bhasin, A. (2013). Credit card dataset for clustering. Accessed: September 2024.

[Huang et al., 2023] Huang, X., Liu, A., and Wang, J. (2023). Energy consumption optimization in data centers using fungi-inspired algorithms. *Journal of Sustainable Computing*.

[Lin et al., 2021] Lin, S., Zhang, Y., and Zhao, H. (2021). Financial data analysis and consumer behavior using fungi kingdom expansion algorithm. *Journal of Financial Data Analytics*.

[Zhao et al., 2022] Zhao, Y., Shen, L., and Lü, Z. (2022). Fungi-inspired algorithms for smart manufacturing scheduling and resource optimization. *Journal of Industrial Optimization*.