POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

EXECUTIVE SUMMARY OF THE THESIS

Implementation of an Agent-Based Recommendation System Using Time-Variant Markov Chains: Investor Personas Forecasting with a Business Case Application

TESI MAGISTRALE IN INGEGNERIA MATEMATICA - MATHEMATICAL ENGINEERING

AUTHOR: LORENZO GROSSI

ADVISOR: DANIELE MARAZZINA

CO-ADVISOR: RAFFAELE ZENTI

ACADEMIC YEAR: 2023-2024

Introduction

The financial world is currently undergoing a revolutionary transformation, driven by factors that are significantly reshaping the way investments are approached. Two major trends, already visible and expected to gain importance in the coming years are: firstly, an aging population[1], which brings a growing demand for quality of life in old age, requiring tailored financial products to meet their expectations; secondly, a drastic shift among younger generations towards non-traditional investment products[2], such as cryptocurrencies and sustainable assets.

To address these evolving challenges, **agent-based** modeling emerges as a fundamental yet underexplored tool for forecasting client behavior, particularly in predicting how the composition of the client base will change over time. This approach segments datasets into **personas** (concrete and identifiable representations of different client groups) and uses evolutionary models to track how each persona's population evolves.

This thesis aims to illustrate the potential of these tools (that can not only be employed in financial applications, but also prove effective in any context involving persona segmentation) through a case study on clients of a financial institution.

Various evolutionary models are introduced, and using Monte Carlo simulations, the investor population is projected over 10 years. Finally, an operational strategy is proposed for the institution to address the observed outcomes, offering a practical application of the forecasting approach.

Our analysis serves as a showcase for the adaptability and flexibility of the method. While addressing a concrete and specific problem, it also aims to inspire others to apply this approach in different contexts, given its potential and the fact that it remains relatively underexplored.

Data and Recommendation System

Our case study is based on a dataset of 5,000 clients from an Italian financial institution. The data includes numerical variables, such as income, wealth, and savings, along with categorical variables that describe each client's socioeconomic attributes. A key feature is the Investor Type classification, which divides clients into clients linked to lump-sum investments, income investors and non-investors.

Our analysis builds upon an agent-based recommendation system developed by Veronica Lucchetti in her thesis, *"Data-Driven Customer Segmentation: A Needs-Based Cluster Analysis for Optimizing Financial Product Recommendations"*[3].

In her work, Lucchetti defines a metric to calculate the distance between data points and, through clustering techniques, identifies six personas (i.e., the resulting clusters). These personas are primarily distinguished by age and Investor Type: Clusters 1 and 2 are linked to lump-sum investments, Clusters 3 and 4 to capital accumulation, and Clusters 5 and 6 are composed of non-investors. Within each pair, one cluster represents younger clients, while the other represents older ones.
An agent-based recommendation system is then implemented, where the agents are the personas identified in Lucchetti's analysis. For each investor client, suitable financial products are recommended, while no products are suggested for non-investors in Clusters 5 and 6.

Models Presentation

This and the following section are the core of our study, where we effectively implement and observe the results of the forecasting models. They also represent the real innovation, as no similar study currently exists in the state of the art.

To project the population of each cluster over the next decade, three different models are defined to simulate the evolution of clients. The purpose of presenting multiple models is to demonstrate the flexibility and adaptability of this forecasting technique, and also to compare the outcomes. All models are based on applying a Gaussian perturbation to each client's numerical features, while categorical features remain unchanged. Additionally, a "death cluster" and a "storage cluster" (for clients entering the client base) are introduced alongside the investor clusters; the time span is 10 years.

Each model follows the algorithm outlined below:

Algorithm 1 Client Persona Identification for Each Year

```
for years from 1 to 10 do
    increment clients' age by 1
    for each client do
        if client died the previous year then
            assign the client to the death cluster for
            the current year
        else
            modify client's features (according to the model)
            simulate client death through a random variable
        end if
    end for
    compute the distance matrix between living clients
    and the original cluster centers
    assign each living client to the cluster with the
    nearest center
    add 200 new clients from the storage cluster
    to the current client base
end for
```

In all models, the distance matrix is computed using the metric introduced by Lucchetti.
As time progresses, the weight assigned to dissimilarities in the investor type feature decreases, reflecting the diminishing importance of the investor type a client was at year 0 as time moves away from that point.

Models Results

Each time a model is simulated, the results produce a distribution of clients among personas from year 1 to year 10. By performing a Monte Carlo simulation of each model with 100 trials, we obtain estimates of the transition probabilities between clusters at different years, which can compose a time-variant Markov Chains.

A time-variant Markov Chain for a given year y is represented by a transition matrix P_y, where the probability of moving from state i to state j between years $y-1$ and y is $P_{i,j|y}$. The transition matrix $P_{0\to n}$ for an n-year jump from year 0 is computed as:

$$P_{0\to n} = P_1 \times P_2 \times \ldots \times P_n$$

The distribution of clients π_n at year n, given the initial distribution π_0 at year 0, is:

$$\pi_n = \pi_0 \times P_{0\to n}$$

In our case study, the initial distribution π_0 of investors is given by:

$$\pi_0 = [737, 759, 804, 1386, 770, 544]$$

The three different models yield very similar outcomes, as confirmed by matrix similarity indices. However, there are significant differences between the matrices for early years and those for later years. In the early years, the matrices show little to no possibility for clients to change investor type, while in later years, there is much greater variability and freedom.

To illustrate the quantities involved, the transition matrices $P_{0\to 5}$ and $P_{0\to 10}$ of the second model, for instance, are shown in Figure 1.

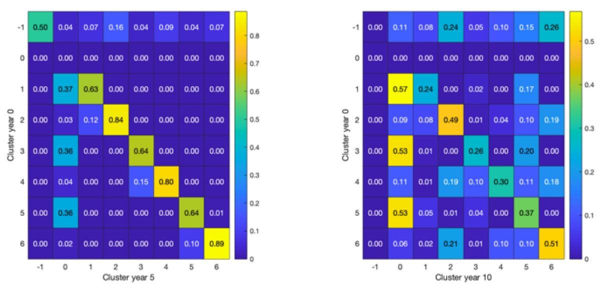

Figure 1: $P_{0\to 5}$ and $P_{0\to 10}$ resulting from the simulations of the second model

The main result observed from the forecasting of client distribution (Figure 2), particularly for years 5 and 10, is an increase in the number of non-investor clients, and a decrease in investors. This issue will be addressed in the following section, where we present a concrete operational strategy that the institution can implement to respond to these forecasts, demonstrating the effective utility of these simulations.

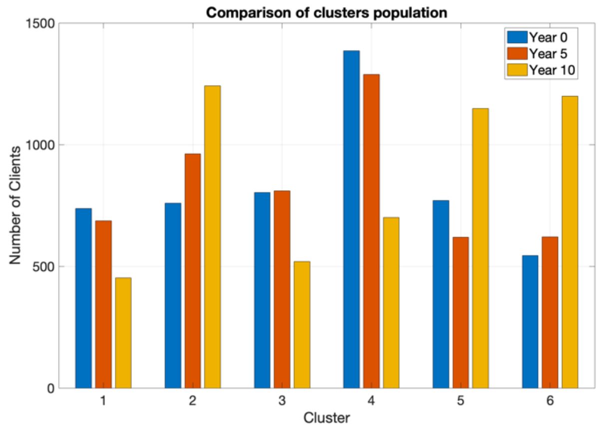

Figure 2: Clients distribution at years 0, 5 and 10

Operational Strategy

To illustrate the usefulness and power of this methodological approach, we present a concrete example that directly impacts the business case, based on the forecasts obtained in the previous section. While this is just one of many possible applications of our model, we believe it is crucial to demonstrate its potential.

In the context of our business study, the predicted increase in non-investor clusters (clusters 5 and 6) and the decrease in investor clusters (clusters 1, 2, 3, and 4) lead to an estimated decline in financial product sales. It's worth noting that in Lucchetti's recommendation system[2], products were suggested only for financial clusters.

We propose an operational strategy that involves maintaining the same recommendation system for investor clusters while suggesting highly specialized financial products for clusters 5 and 6 to address their needs and avoid financial risks to which they are averse. Through a decision-tree-like segmentation of these clusters, we identify specific categories of people and match them with appropriate insurance products.

The details of these recommendations and the identified subclusters are presented in Figure 3 and Figure 4.

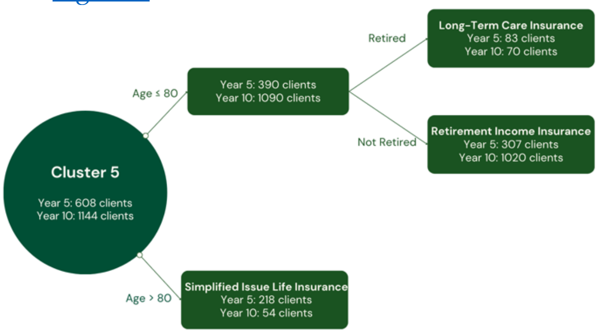

Figure 3: Segmentation of cluster 5 and insurance products suggested

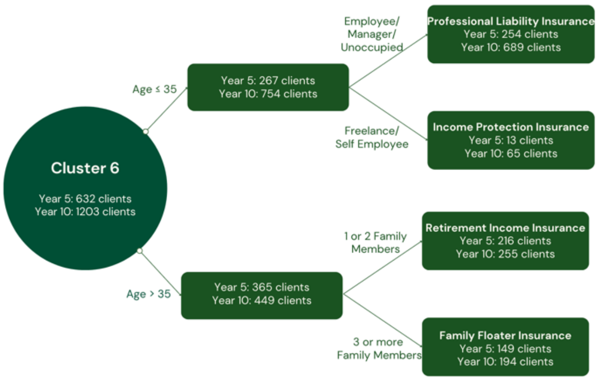

Figure 4: Segmentation of cluster 6 and insurance products suggested

To estimate the total number of products sold in different years (which provides a rough but important estimate), research indicates that insurance products will be purchased by approximately 40% of the individuals to whom they are suggested. Figure 5 presents the results based on this estimate, showing an increase in total products sold of 5.1% at year 5 compared to year 0, and of 4.6% at year 10.

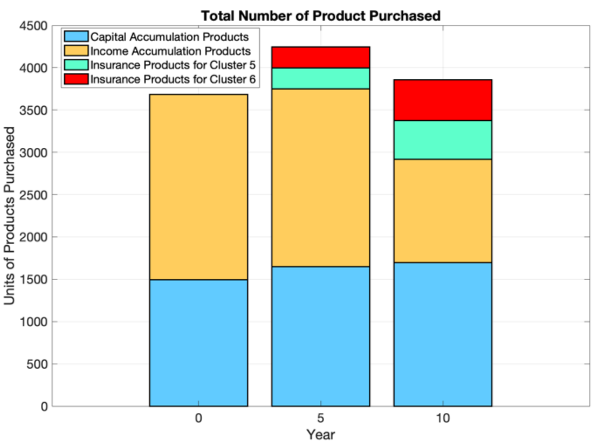

Figure 5: Total number of estimated sales

Discussions

The aim of this work is not to claim the ability to predict the future of a financial institution, but rather to create a plausible, real-world agent-based scenario illustrating how client distribution may evolve. To demonstrate the utility of this model, we applied it to one of its many potential applications: making informed decisions to address the evolving client base indicated by the model and suggesting potential future directions, particularly in forecasting client needs and identifying suitable financial or insurance products.

Considering the outcomes, we can confidently assert that the study's primary objective has been achieved: developing a flexible system that, in synergy with expert insights and intuition, can serve as a critical tool for industries to navigate changes and plan for the future.

Key Findings

The model developed, with minimal recommendations and adjustments, is versatile and applicable to a wide range of scenarios beyond the financial sector. Its core strength lies in making the fewest assumptions possible:

- Availability of a reliable dataset, where a quantity of interest (e.g., Investor type) is defined for each client, along with other features describing the client base.
- Clients can be categorized into a small number of personas, meaning individuals can be grouped based on similar needs.
- The ability to model the evolution of clients over time, driven by knowledge and rational assumptions.

In our specific case study, the main results of the simulations are:

- A significant increase in non-investor clients, alongside a notable decline in investors, leading to a corresponding reduction in financial product sales.
- Forecasts and operational strategies effectively address the main challenges discussed in the Introduction, including the need for a wide range of products

tailored to older clients and the significant differences in product demands between younger and older individuals.

Implications

The goal of this work is not just to create a financial evolutionary scenario showing how investors change over time, but to develop a broader model that, with minimal adjustments, could serve as a foundational tool in various contexts involving the evolution of people and the categorization of personas. This model could be applied to numerous scenarios, ranging from patient management in healthcare to client profiling, student classification, and beyond.

Our work introduces an innovative approach by focusing on real-world future scenarios of investors, a topic not yet explored in the current state of the art. The model's scalability allows it to be applied to a wide range of clients, including those from diverse socioeconomic backgrounds, international markets, and larger datasets. Its flexible structure also makes it easily integrable with other predictive technologies.

Recommendations

The decision to use a time-variant Markov Chain was driven by its adaptability, which surpasses that of homogeneous models. The time-varying chains effectively capture changes across different years, validating this approach. In addition to this flexibility, the model allows for greater freedom in the elaboration of evolution techniques.

To fully leverage the potential of the model, we offer two key recommendations:

- **Strong Persona Definition**: Ensure that the entire framework is grounded in a robust definition and profiling of personas, as this serves as the foundation of the model.
- **Models Testing**: When evaluating the evolutionary model, it is essential to test the models across multiple trials. Running extreme trials can provide valuable insights into the system's dynamics.

Conclusion

This study successfully demonstrates, through a concrete case study, the potential of Time-Variant Markov Chains to forecast how the client base of a financial institution can evolve over a given period and how to address these changes. Based on a reliable algorithm for implementing agent-based recommendations, three different models of client feature variations have been developed to explore the model's flexibility and compare the resulting outcomes.

The results indicate an increase in non-investor clients and a decrease in investors, leading to a reduction in financial product sales. To address these findings and demonstrate the practical application of the model, we have proposed a future business plan for the institution, which includes introducing insurance products tailored to specific subclusters within the non-investor groups.

This work not only provides a concrete example of a plausible future scenario and its implications but also highlights how these models, requiring only evolutionary rules and a suitable clustering technique, can be extended beyond the financial sector to any context involving agent-based recommendation systems. The model's adaptability and potential are its core strengths, making it a potentially winning tool for developing successful strategies in any field involving agent evolution, of which there are countless examples.

Bibliography

[1] Citi. Research report, 2024.
Electronic document available at https://www.citifirst.com.hk/home/upload/citi_research/rsch_pdf_30259888.pdf

[2] Bank of America. BofA private bank study of wealthy Americans finds generational shift in attitudes towards wealth, June 2024.
Electronic document available at https://newsroom.bankofamerica.com/content/newsroom/press-releases/2024/06/bofa-private-bank-study-of-wealthy-americans-finds-generational-.html

[3] V. Lucchetti. Data-driven customer segmentation: A needs-based cluster analysis for optimizing financial product recommendations. Unpublished master's thesis, 2023-2024.

Acknowledgments

I would like to thank Professor Marazzina for being my thesis supervisor and Veronica for all the work we did together, her endless support, and the positivity she has always brought me.
I would also like to thank Professor Zenti for generously sharing his deep knowledge and for showing me, in a very concrete way, a precise and sharp working method that I will always carry with me as an example. I also thank him for the confidence he instilled in me during the moments when I needed it most; I am proud of the thesis we produced, and much of the credit goes to him.