# Experimental Report on the Comparison Between Generalized Mapping and Markov Chain in E-commerce Fraud Detection

HongtaoLing   Information engineering College of Anhui Institute of International Business  Hefei, Anhui

## Abstract

This experiment focuses on the field of e-commerce fraud detection, aiming to compare the performance of the generalized mapping rule system and the Markov chain state transition system. By programming with Python, 5000 simulated order data with 9 features were generated, and the two systems were used for processing respectively. The experimental results were analyzed from aspects such as time complexity and accuracy to provide data support for the selection of e-commerce risk control strategies. It was found that the generalized mapping has a relatively high accuracy but insufficient recall for fraud orders, while the Markov chain has high processing efficiency but serious misdetection.

**Keywords:** Generalized mapping; Generalized functions; Markov Chain; E-commerce Fraud Detection

## 1. Introduction

Generalized mapping, as a powerful analysis tool, shows unique advantages in many fields. Its second form can construct a unified structural pattern and present different mapping states according to parameter changes. In the e-commerce fraud detection scenario involved in this experiment, the generalized mapping identifies fraud orders by integrating multiple feature rules and using a weighted judgment method. Compared with the traditional single-rule judgment, this mode can comprehensively consider various factors and effectively improve the comprehensiveness and accuracy of detection, thus safeguarding the secure transactions of e-commerce platforms.

## 2. Experimental Content

### 2.1 Data Generation

The numpy library of Python is used to generate 5000 simulated e-commerce order data. Each order contains 9 features, such as the number of items (ranging from 1 to 20), order value (from 10 to 1000), user level (bronze, silver, gold), user age (from 18 to 70), monthly purchase frequency (from 0.1 to 10.0), historical return rate (from 0 to 0.5), device type (mobile, desktop, tablet), location risk (from 0 to 1), and IP risk score (from 0 to 1). At the same time, a real fraud label (is_fraud) is generated for each order according to complex fraud rules. For example, the fraud probability is affected by multiple features. Situations such as the number of items greater than 15 and the order value less than 100 will increase the fraud probability. Finally, whether it is a fraud order is determined by comparing a random number with the fraud probability.

### 2.2 Implementation of Generalized Mapping Risk Control System

A rule-based generalized mapping risk control system is constructed. The RiskRule base class is defined, which includes the rule name and weight. Multiple subclasses implement specific rule judgment logic, such as QuantityRule to judge whether the number of items is greater than 15, and ValueRule to judge whether the order value is less than 100. In the RiskSystem class, all rules are integrated. For each order, each rule is checked in turn to see if it is triggered, and the weights of the triggered rules are accumulated to get the risk score. If the score exceeds the threshold of 0.5, the order is determined to be a high-risk (fraudulent) order.

## 2.3 Implementation of Markov Chain Risk Control System

A Markov chain risk control system based on state transition is designed. A state transition diagram is predefined. The system starts from the start state, goes through states such as check_quantity and check_value in turn, and finally transfers to the low_risk (normal) or high_risk (fraudulent) state. A preset transition probability is set for each check state, and the transition probability is dynamically adjusted according to the order features. For example, if the number of items in an order is greater than 15 in the check_quantity state, it will directly transfer to the high_risk state with a 100% probability, so as to simulate the risk control logic of step-by-step in-depth inspection.

# 3. Experimental Process

1. **Data Initialization**: Run the generate_orders function to generate 5000 simulated order data and extract the real fraud label y_true.

2. **Performance Test**: Set the test order scales to 100, 300, 1000, 3000, and 5000, and conduct processing time tests on the generalized mapping and Markov chain systems respectively. For each scale of order subset, record the total processing time of the generalized mapping system for each order and the total processing time of the Markov chain system for orders, and compare their time complexities.

3. **Accuracy Verification**: Use the generalized mapping system and the Markov chain system to predict fraud for all 5000 orders respectively to obtain the prediction labels y_pred_gm and y_pred_mk. Compare the prediction labels with the real label y_true, use accuracy_score to calculate the accuracy, and generate detailed classification indicators such as precision, recall, and F1 score through classification_report.

4. **Result Display**: Organize the time data of the performance test and the classification indicator data of the accuracy verification, and generate a processing time comparison chart and an accuracy comparison chart through the plot_results function to intuitively display the performance of the two systems in different aspects. At the same time, output the rule triggering details of a specific order (such as order ID 0) in the generalized mapping system to assist in understanding the system decision-making process.

# 4. Result Analysis

## 4.1 Time Complexity

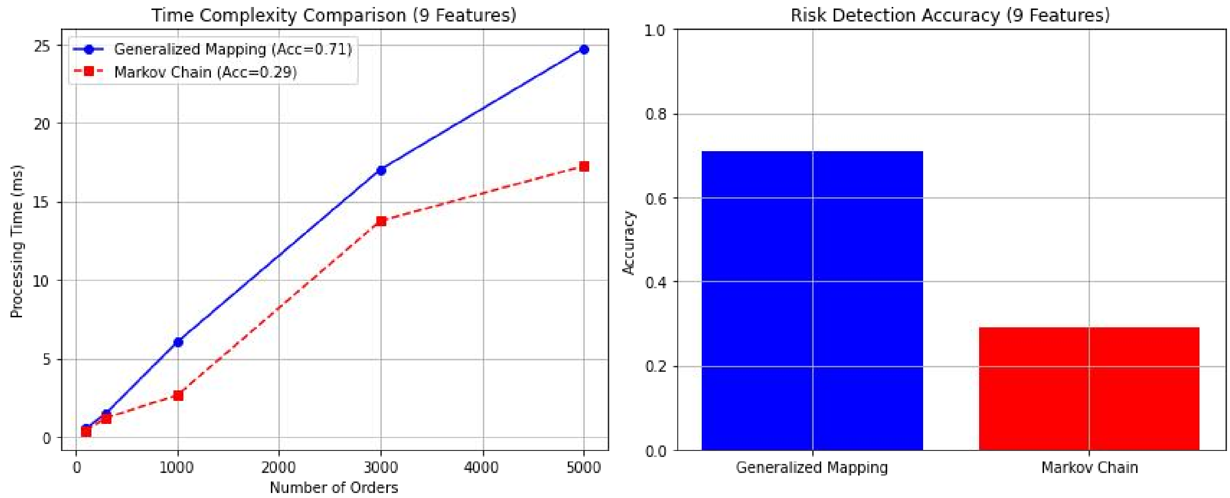
As the number of orders increases, the processing time of both the generalized mapping and Markov chain systems increases. However, under each test scale, the processing time of the Markov chain is shorter than that of the generalized mapping. For example, when the test scale is 1000, the time for generalized mapping is 6.03 ms, and that for Markov chain is 2.64 ms. This shows that the Markov chain has an advantage in processing efficiency and a relatively lower time complexity.

## 4.2 Accuracy

The accuracy of generalized mapping is 0.71, and that of Markov chain is 0.29. The generalized mapping is significantly higher than the Markov chain. From the details of the classification report, the generalized mapping has a strong ability to identify non-fraudulent orders (the recall rate of the "False" class is 0.99), but it is insufficient in identifying fraud orders (the recall rate of the "True" class is only 0.03); the Markov chain can identify most fraud orders (the recall rate of the "True" class is 1.00), but misclassifies a large number of non-fraudulent orders as fraudulent (both the precision and recall rates of the "False" class are 0.00).

Link to the experimental code:https://gitee.com/riririiririiriir/First-Study-on-Generalized-Mapping-and-Markov/tree/master/

The experimental result graph is as follows：




# 5. Conclusion

This experiment compares the application of generalized mapping and Markov chain in e-commerce fraud detection. The generalized mapping performs better in accuracy and can effectively identify non-fraudulent orders, but there is a lack of recall for fraud orders; although the Markov chain has high processing efficiency, the problem of misdetection is serious, resulting in low overall accuracy. At present, the research on the combination of generalized mapping and Markov chain in scenarios such as e-commerce risk control is still shallow. It is hoped that colleagues will explore together to further tap the potential of generalized mapping in Markov-related fields, optimize the algorithm combination, improve the comprehensive performance of practical applications such as e-commerce fraud detection, and promote the technological development and innovation in this field.

# References

[1] Ling H. Generalized Mapping Theory — Used to Describe Phenomena That Cannot Be Characterized by Generalized Functions. Preprints 2025, 2025080640. https://doi.org/10.20944/preprints202508.0640.v1