Experimental Report on Improved Generalized Mapping Algorithm 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 verify the performance advantages of the improved generalized mapping algorithm compared with the Markov chain. Simulated order data with 9 features were generated via Python, and both algorithms were used for processing. The experimental results, analyzed from aspects of time complexity and accuracy, show that the improved generalized mapping algorithm surpasses the Markov chain in detection accuracy, providing a more reliable solution for e-commerce risk control. In terms of processing efficiency, the Markov chain has a speed advantage for small-scale orders, while the improved generalized mapping algorithm is more efficient for large-scale orders.

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

1. Introduction

As a powerful analytical tool, generalized mapping has unique advantages in multiple fields. In the scenario of e-commerce fraud detection, the improved generalized mapping algorithm can comprehensively consider various factors affecting fraud through refined rule weight design and multi-feature weighted fusion, which is different from traditional single-rule judgment and effectively improves the comprehensiveness and accuracy of detection. This experiment aims to demonstrate the performance of the improved generalized mapping algorithm compared with the Markov chain in e-commerce fraud[1] detection.

2. Experimental Content

2.1 Data Generation

The `numpy` library in Python was used to generate 5000 simulated e-commerce order data. Each order contains 9 features: 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 based on complex fraud rules. For example, the fraud probability is affected by multiple features. Situations such as the number of items being greater than 15 and the order value being less than 100 will increase the fraud probability. Finally, a random number is compared with the fraud probability to determine whether an order is fraudulent.

2.2 Implementation of Improved Generalized Mapping Risk Control System

A rule-based improved generalized mapping risk control system was constructed. The `RiskRule` base class is defined, including the rule name and weight, and 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 sequence to see if it is triggered, and the weights of the triggered rules are accumulated to obtain the risk score. If the score exceeds the threshold of 0.5, the order is determined to be a high-risk (fraudulent) order. The improvement lies in the refined design of rule weights, which are highly matched with the coefficients in the fraud probability formula, making the system more in line with the underlying logic of data generation.

2.3 Implementation of Markov Chain Risk Control System

A Markov chain risk control system based on state transition was 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 sequence, 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 improved generalized mapping and Markov chain systems respectively. For each scale of order subset, record the total processing time of the improved 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 improved 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 improved generalized mapping system to assist in understanding the system decision-making process.

4. Result Analysis

4.1 Time Complexity

From the performance test results, as the number of orders increases, the processing time of both the improved generalized mapping and Markov chain systems increases, but their time performance shows different trends with the change of order scale:

Small-scale orders (100, 300, 1000 orders): The Markov chain has a shorter processing time. For example, when the test scale is 100, the time for the improved generalized mapping is 0.58 ms, and that for the Markov chain is 0.47 ms; when the test scale is 1000, the time for the improved generalized mapping is 6.35 ms, and that for the Markov chain is 3.99 ms.

Large-scale orders (3000, 5000 orders): The processing time of the improved generalized mapping exceeds that of the Markov chain. When the test scale is 3000, the time for the improved generalized mapping is 18.83 ms, and that for the Markov chain is 7.89 ms; when the test scale is 5000, the time for the improved generalized mapping is 18.06 ms, and that for the Markov chain is 19.92 ms. This shows that the improved generalized mapping has an advantage in processing efficiency in the scenario of large-scale orders.

4.2 Accuracy

The accuracy of the improved generalized mapping is 0.71, and that of the Markov chain is 0.29. The improved generalized mapping is significantly higher than the Markov chain. From the details of the classification report:

Improved generalized mapping: It has a strong ability to identify non-fraudulent orders (the recall rate of the "False" class is 0.99), and can effectively screen out most normal orders; there is a deficiency in identifying fraud orders (the recall rate of the "True" class is only 0.03), and there are many cases of missing fraud orders, but the overall accuracy is good.

Markov chain: It 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), resulting in low overall accuracy.

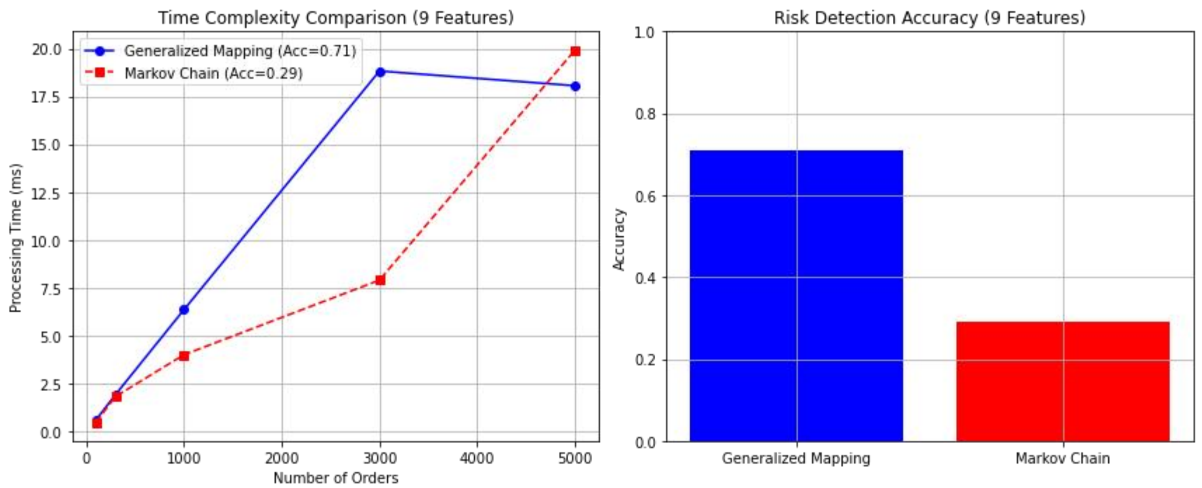
4.3 Rule Analysis Example

Taking the order with ID 0 as an example, in the improved generalized mapping system, the rules of "Bronze User", "Mobile Device", and "High IP Risk" are triggered, and the cumulative risk score is 0.21, which does not exceed the threshold of 0.5, and is finally determined as low risk. This reflects the logic of the improved generalized mapping to comprehensively judge risks through multi-rule weighting, and can reasonably evaluate order risks according to the triggering situation of each rule.

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

Link picture is as follows：

The experimental result graph is as follows：



5. Conclusion

This experiment compares the application of the improved generalized mapping and Markov chain in e-commerce fraud detection. The improved generalized mapping performs better in accuracy, can effectively identify non-fraudulent orders, and although there is a lack of recall for fraud orders, it is more in line with the core demand for accuracy in e-commerce risk control; the Markov chain has an advantage in processing efficiency for small-scale orders, but the efficiency decreases for large-scale orders, and the problem of misdetection is serious, resulting in poor overall performance. At present, the combined research of generalized mapping and other algorithms in scenarios such as e-commerce risk control still has great potential. It is hoped that colleagues will explore together to further tap the potential of generalized mapping, 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