Journal of Artificial Intelligence Practice (2024)
Clausius Scientific Press, Canada
DOI: 10.23977/jaip.2024.070119
ISSN 2371-8412 Vol. 7 Num. 1

Rule-based Matching and Hidden Markov Model-based Warning for Brushing Behavior

Jian Wang, Yang Liu

Shandong Yantai Tobacco Co., Ltd., Yantai, 264000, China

Keywords: Rule-based matching, Hidden Markov Model, Brushing behavior, Warning system, Tobacco industry, Consumer behavior, Point redemption

Abstract: This paper proposes a warning system for brushing behavior based on the tobacco industry's reward activities by combining rule-based matching and Hidden Markov Models. Targeting behaviors where customers make large-scale purchases during promotional events that deviate from their individual historical consumption patterns, the system employs rule-based matching for initial assessment, followed by a more in-depth behavioral analysis using Hidden Markov Models. Experimental results demonstrate the significant effectiveness of the system in warning against brushing behavior, providing an effective monitoring tool for the tobacco industry and similar sectors.

1. Introduction

With the rapid development of e-commerce, brushing behavior has brought significant economic losses to businesses. In order to effectively monitor and prevent brushing behavior, this paper proposes a warning system that combines rule-based matching and Hidden Markov Models. Specifically, we focus on brushing behavior during the tobacco industry's reward activities, as some customers may rapidly accumulate points through bulk purchases during these events and then redeem prizes in a short period. This behavior deviates from their individual previous consumption habits, highlighting the need for an efficient warning system to identify and prevent such brushing activities.

2. Challenges and Responses to Brushing Behavior in the E-commerce Era

With the flourishing development of e-commerce, brushing behavior has gradually emerged as a significant threat to the economic well-being of businesses. This chapter will delve into the menace that brushing behavior poses to the economic landscape of enterprises and explore the reasons behind the widespread concern about the issue in contemporary society.[1]

2.1. Challenges of Brushing Behavior in the E-commerce Era

The rise of e-commerce has brought convenience to both merchants and consumers, but brushing behavior has posed economic challenges. Brushing refers to consumers engaging in fraudulent transactions through fabricated or deceptive means to gain undue benefits. Brushing behavior manifests in various forms, including false transactions and malicious reviews. Fabricated transaction activities not only deceive the platform and merchants but also mislead other consumers in their shopping decisions, disrupting the fairness of market competition. Simultaneously, brushing behavior that circumvents the platform's monitoring system through technological means has inflicted significant economic losses on e-commerce enterprises. This undermines the profitability of merchants and diminishes the overall vitality of the entire e-commerce market.

To effectively address this issue, the e-commerce industry urgently requires a comprehensive and intelligent mechanism for monitoring and preventing brushing behavior. Only through innovative technological means can we establish a timely detection and counteraction mechanism against brushing behavior, thereby safeguarding the healthy ecosystem of the e-commerce industry and ensuring the legitimate rights and interests of both merchants and consumers.[2]

2.2. Brushing Challenges Arising from Tobacco Industry Reward Activities

The tobacco industry, as a distinctive sector, faces significant challenges of brushing behavior during reward activities. During these events, some consumers engage in unusually large-scale purchases of tobacco products to rapidly accumulate points and redeem prizes within a short period. This behavior not only deviates from individual normal consumption patterns but also poses a serious threat to the promotional strategies of tobacco enterprises.

Brushing behavior exhibits unique characteristics in reward activities within the tobacco industry, rendering traditional monitoring methods insufficient. To address this issue, the research team is dedicated to constructing a specialized brushing detection system tailored to the tobacco industry, building upon e-commerce foundations. This system will integrate rule-based matching and Hidden Markov Models, providing a more comprehensive and intelligent brushing detection solution through in-depth analysis of purchasing patterns and point acquisition behaviors.[3]

3. Rules for Identifying Brushing Behavior

This chapter delves into the methods for initially identifying brushing behavior through specific rules. Firstly, we introduce the rule-matching approach, swiftly and accurately identifying potential brushing activities based on conditions such as exceeding purchase thresholds and deviations in consumption behavior from historical patterns.

3.1. Rule Matching

Rule matching is a crucial step in identifying brushing behavior, involving a series of explicit conditions tailored to screen customer actions. This step primarily focuses on identifying brushing characteristics, such as purchase quantity, frequency, and point accumulation. Particularly in reward activities within the tobacco industry, we swiftly discern abnormal behavior through specific rules.

3.1.1. Rules for Purchase Quantity and Frequency

In the rule-matching process, the purchase quantity and frequency are critical factors for judgment.[4]By establishing thresholds for purchase quantity and frequency, this preliminary assessment assists in identifying customers who make multiple purchases in a short period, thereby enhancing the accuracy of identification.

3.1.2. Rule for Consistency in Consumption Patterns

To comprehensively determine brush-order behavior, customer consumption pattern analysis is introduced. By monitoring the magnitude of changes in user consumption patterns, abnormal behaviors are detected and flagged, providing a basis for more detailed analysis.

The establishment of these rules allows us to quickly and accurately identify potential brush-order activities in the rule-matching phase. However, as a preliminary assessment stage, rule matching needs to be combined with in-depth analysis to ensure the precise identification and meticulous handling of brush-order behavior. The following sections will introduce more in-depth analysis methods to build a more comprehensive brush-order early warning system.[5]

3.2. Brushing Characteristics in the Tobacco Industry Reward Activities

During reward activities in the tobacco industry, brushing behaviors exhibit unique characteristics that require in-depth analysis to construct a comprehensive brushing model. This section focuses on analyzing the distribution of purchase times and the speed of point accumulation to accurately identify brushing.

3.2.1. Purchase Behavior Distribution

Brushing in the tobacco industry is highly likely to occur during reward activities, and the behavior, such as the time distribution and purchase quantity distribution, differs from normal purchases. By analyzing purchase behaviors, examining data such as the distribution of purchase times and time density, a preliminary judgment can be formed.

3.2.2. Point Growth Behavior

One purpose of brushing is to rapidly accumulate points for exchanging prizes. Under normal circumstances, point accumulation occurs at a stable speed. Brushing may manifest as quickly accumulating a large number of points within a short period. By setting point speed rules, abnormal brushing with high speed can be identified, providing clues for analysis.[6]

3.3. In-Depth Analysis of Brushing Behavior

After rule matching, in-depth analysis of brushing behavior is crucial to comprehensively understand the patterns and characteristics of brushing. At this stage, through mining purchase and account behaviors, the authenticity and complexity of brushing are confirmed.

3.3.1. Purchase Behavior Patterns

In-depth analysis of purchase patterns helps identify covert and complex brushing methods. Examining details such as purchase times, types of products, amounts, etc., aids in identifying potential brushing. For example, brushing might occur during peak periods of normal purchases to obscure its behavior. Refining rules can capture these behaviors, strengthening the warning system.

3.3.2. Device Anomaly Identification

Considering device information, such as type and switching frequency, enhances detection capabilities. Brushers may operate multiple devices to avoid monitoring on a single device. Establishing rules to identify frequent device switching increases accuracy.

In-depth analysis of purchase, account, and device behaviors helps understand the complexity and diversity of brushing. This stage provides detailed feature information for the final warning system, creating an intelligent and reliable monitoring solution.

4. Hidden Markov Model Analysis

This chapter explores how to construct a Hidden Markov Model (HMM) based on the initial judgment from rule matching. It provides detailed insights into the model's construction, training, and optimization processes, aiming to enhance the accurate identification of brushing behavior.

Hidden Markov Model (HMM) is a dual stochastic process, one being a Markov process, and the other being an explicit stochastic function set.For a discrete Hidden Markov Model, let Q represent the hidden states of a discrete-time finite-state, one-step homogeneous Markov chain., q_t Here, t represents the hidden states of this Hidden Markov Model at time .Let O be an observable general random process. O_t Where t is the observed value of this hidden Markov model at time.

HMM is composed of a quintuple, generally represented as

$$\lambda = (S, V, A, B, \Pi):$$

S is a set of hidden states containing N states, represented as: $S = (s_1, s_2, \cdots, s_N)$;

V is a set containing M observable outcomes, represented as

$$V = (v_1, v_2, \cdots, v_M);$$

A It is the matrix of hidden state transition probabilities, $A = [a_{ij}]_{N \times N}$,These two $a_{ij} = P(i_{t+1} = q_j | i_t = q_i), i = 1, 2, \cdots N; j = 1, 2, \cdots N$ at time t t in state q_i under the condition at time $t+1$ transition to state q_j the probability of.

B is the observation probability matrix, $B = [b_j(k)]_{N \times M}$,These two $b_j(k) = P(o_t = v_k | i_t = q_j), k = 1, 2, \cdots M; j = 1, 2, \cdots N$ is at time t in state q_j transitioning to state. v_k the probability of.

Π is the initial state probability vector., $\Pi = (\Pi_i)$,These two $\Pi_i = P(i_1 = q_i), i = 1, 2, \cdots N$ at time $t = 1$n state q_i the probability of.

4.1. Model Construction

In the model construction phase, we will design a Hidden Markov Model (HMM) based on the preliminary determination results from rule matching to more comprehensively and intelligently identify fraudulent activities. Considering that fraudulent behaviors may have multiple states, we will establish an HMM that includes different states, such as fraudulent states and normal purchase states. The construction of the model will focus on the state transition probabilities and observation probabilities of fraudulent activities.

4.1.1. State Definition

In the state definition phase, we categorize customer purchasing behaviors into different states based on the preliminary determinations from rule matching. These states may include normal purchase states, suspected fraudulent states, etc. That is, define the sets S and V in .At the same time, calculate the initial state probability vector Π under this definition. For example, customers whose purchase quantity exceeds the threshold may be classified into a suspected brush order state. The explicit definition of these states lays the foundation for the state space of the Hidden Markov Model, enabling the model to better understand and capture the characteristics of different purchasing behaviors.

4.1.2. Transition Probability

Subsequently, we will establish the transition probability matrix, describing the transition process of brush order behavior between different states. By analyzing historical data, we can estimate and construct the transition probability matrix A, namely constructing the matrix for hidden state transitions. For example, if a customer's previous state is a normal purchasing state, the next state may continue to remain a normal purchase, or it may transition to a suspected brush order state. The hidden state transition probability matrix A reflects the dynamic evolution process of brush order behavior.

4.1.3. Observation Probability

To more accurately observe and assess customers' actual purchasing behavior, we introduce observation probability. The observation probability represents the probability of observing different values under a specific state, indicating the characteristics of customer purchasing behavior in a given state, i.e., constructing the observation probability matrix B in HMM. Through the analysis and statistics of historical data, we can establish the probability distribution of observed values such as purchase quantity and purchase frequency under different states. These observation probabilities will contribute to enhancing the model's accuracy in identifying brush order behavior.

Through the above steps, we have successfully established a Hidden Markov Model (HMM) that includes state definition, state transition probabilities, and observation probabilities. This model will provide a solid theoretical foundation for subsequent training and optimization, enabling more accurate identification and prediction of brush order behavior.

4.2. Model Training and Optimization

After the construction of the Hidden Markov Model (HMM), to enhance its accuracy in identifying brush order behavior, we will proceed with model training and optimization. This process involves utilizing historical data, adjusting model parameters, and continuously validating the model's predictive results.

4.2.1. Model Training

Model training is a crucial stage where, by using historical data, the model gradually learns and adjusts its internal state transition probabilities and observation probabilities to better reflect the actual distribution of brush order behavior. The training process employs iterative algorithms, with the model continuously comparing itself to historical data to incrementally improve sensitivity and accuracy in identifying brush order behavior.

Model training primarily involves aspects such as collecting historical data, initializing parameters, and iterative training. The model predicts based on the training set and adjusts according to observed outcomes. This process iterates until the model's performance reaches a satisfactory level.

4.2.2. Model Optimization

Upon completing model training, we proceed with model optimization, continually adjusting its parameters by comparing predicted results with actual observations to enhance its accuracy in identifying brush order behavior.

During model optimization, adjustments to the model's state transition probabilities and observation probabilities are made to better align with real-world scenarios. This process requires careful execution to prevent issues of overfitting or underfitting.

Through the iterative process of model training and optimization, our aim is to establish a Hidden Markov Model capable of accurately and robustly identifying brush order behavior in real-world scenarios, providing businesses with a robust brush order alert tool.

4.3. Model Evaluation and Performance Monitoring

In the analysis of the Hidden Markov Model, beyond construction and training, we also need to conduct model evaluation and performance monitoring. This stage aims to validate the model's robustness and generalization capabilities in real-world environments, ensuring its resilience to brush order behavior in different scenarios.

4.3.1. Model Evaluation

Model evaluation is a crucial step in ensuring the superior performance of the model. By using an independent test dataset, we can assess the model's performance on unseen data. Common evaluation metrics include accuracy, recall, precision, etc., providing a comprehensive understanding of the model's performance.

4.3.1.1 Confusion Matrix Analysis

The confusion matrix clearly displays the quantities of true positives, false positives, true negatives, and false negatives, aiding in identifying areas where the model excels and areas that require improvement. Through the confusion matrix (Table 1), the model's accuracy is calculated as 0.717, and the model's recall is 0.238. It can be observed that the model exhibits relatively high accuracy and performs well in identifying brush order behavior. However, manual intervention is still required to determine the true existence of brush order behavior characteristics.

Table 1: Confusion Matrix

	Actual Results		
Model Prediction Results		brushing orders	normal consumption
	brushing orders	285	125
	normal consumption	3	39

4.3.1.2 ROC Curve Analysis

Through the ROC curve, as shown in Figure 1, we can plot the true positive rate and false positive rate at different thresholds to comprehensively evaluate the model's performance. The area under the ROC curve (AUC) is also a key indicator for assessing the model's quality. The preliminary model achieved a good AUC score of 0.762 in the test set.

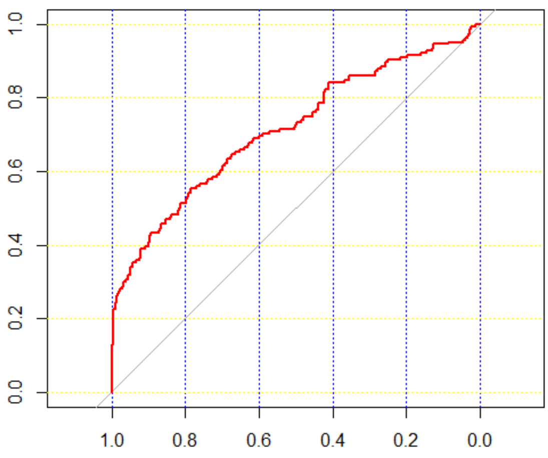

Figure 1: ROC Curve for Model Evaluation

4.3.2. Performance Monitoring

Model performance monitoring is an ongoing process aimed at maintaining the efficiency of the model over time. We will establish a monitoring system to track the model's performance in real-world operations, promptly identifying potential issues.

4.3.2.1 Real-time Data Stream Analysis

Through real-time data stream analysis, we can monitor the model's performance on new data. Timely detection of the model's adaptability to new situations allows us to adjust the model's parameters, ensuring its accuracy in a constantly changing environment.

4.3.2.2 Threshold Adjustment

Adjust the model's output thresholds based on actual operational conditions. This helps balance accuracy and recall, ensuring the model produces appropriate fraud prediction results in different scenarios.

Through model evaluation and performance monitoring, continuously optimize the Hidden Markov Model (HMM) to ensure efficient and accurate identification of fraudulent activities in the dynamic e-commerce environment. This process will provide businesses with a reliable fraud detection system, assisting them in better safeguarding their economic interests.

5. Conclusion

The proposed fraud detection system, based on rule matching and Hidden Markov Models (HMM), has demonstrated significant effectiveness during incentive programs in the tobacco industry. The combination of initial assessment through rule matching and in-depth analysis using HMM allows for more accurate identification of fraudulent activities, providing businesses with an effective means of monitoring and prevention. Future research directions may involve expanding the dataset and optimizing model algorithms to further enhance the system's performance and practicality.

References

[1] Zhou Cong. Longest matching rules in the dual-retrieval corrected domain name system[J]. Computer Engineering and Science. 2023, 45(09): 1572-1577.

[2] Xiao Liping. Research on course teaching quality evaluation based on Hidden Markov Models[J]. Heilongjiang Higher Education Research. 2022, 40(01): 151-155.
[3] Yao Runqi. Unfair competition recognition of online brushing behavior[J]. China Price Supervision and Anti-Monopoly. 2023(03): 58-60.
[4] Peng Renwang. Application of multiple pattern matching algorithm in automatic detection of network intrusion[J]. Automation Technology and Applications. 2021, 40(03): 92-95+101.
[5] Wang Chao. Types, causes, and regulation of brushing behavior[J]. Journal of Fujian Police College. 2020, 34(01): 95-100.
[6] Zhang Xiangya. Research on water quality evaluation of fish ponds based on Hidden Markov Models[J]. Fujian Computer. 2023, 39(04): 67-69.