Development of an Intelligent Recommendation System for Cross-Border E-commerce Platforms using Data Mining Techniques

Hui Zhang

Shandong Vocational College of Science and Technology, Weifang City, Shandong Province 261053, China

39200317@qq.com

Abstract. As cross-border e-commerce grows, intelligent recommendation systems are crucial for platform competitiveness. This research aims to leverage data mining techniques to provide personalized, accurate recommendations. It outlines using data mining in recommendations, and proposes a development process including data preprocessing, feature engineering, model training, and system implementation with a Spark streaming architecture. Key modules are designed for data processing, model training, and online recommendations. Evaluation metrics are established to test performance. The research shows data mining techniques can effectively uncover user interests and patterns for personalized intelligent recommendations. It provides insights into designing cross-border e-commerce recommendation systems. In summary, this research demonstrates data mining techniques can enable effective personalized recommendations, with an outlined development process and insights for system design.

Keywords: Data Mining Techniques; Intelligent Recommendation System; Recommendation System Modeling

1 Introduction

With the growth of cross-border e-commerce, the surge in product choices has led to consumer decision difficulties, impacting purchases. Accurately recommending products to target users has become critical for e-commerce platforms to improve competitiveness. However, the increasing number of products makes it challenging for traditional recommendation systems. Relying solely on rules yields limited results. Therefore, leveraging data mining techniques to provide intelligent and personalized recommendations has become urgent. This research aims to use data mining algorithms to uncover user interest association rules and develop an intelligent recommendation system. We first overview applying data mining in recommendations, then propose a system development process based on data mining techniques. Next, we present a

B. K. Kandel et al. (eds.), *Proceedings of the 2023 3rd International Conference on Business Administration and Data Science (BADS 2023)*, Atlantis Highlights in Computer Sciences 19,
https://doi.org/10.2991/978-94-6463-326-9_36

system design and establish an evaluation mechanism, to offer insights for realizing an intelligent cross-border e-commerce recommendation system [1].

2 Relevant Technology Introduction

2.1 Data Mining Techniques

In recommendation systems, the following categories of data mining techniques are primarily employed:

Classification algorithms like decision trees and Naive Bayes classify users based on demographic features to analyze preferences of different user groups; Clustering techniques like K-means and DBSCAN group users into clusters based on their interests, enabling neighborhood-based collaborative filtering; Association rule learning approaches like Apriori and FP-growth find associations between products and user features to generate recommendations; Deep learning models like neural networks learn complex user representations from large datasets for personalized recommendations [2].

2.2 Intelligent Recommendation Systems

An intelligent recommendation system is a system that utilizes technologies such as data mining and machine learning to provide personalized recommendations based on user's historical behavior and interest features. The recommendation system consists of modules like data collection, data analysis, user modeling, algorithm implementation, and result evaluation. Common recommendation algorithms include content-based filtering, collaborative filtering, and association rules. Recommendation systems can enhance user engagement, improve user experience, and find extensive applications in e-commerce, social media, news, and more. In recent years, the application of deep learning techniques has significantly enhanced the performance of recommendation systems [3].

2.3 Cross-Border E-commerce Platforms

Cross-border e-commerce platforms utilize the Internet and big data technologies to transcend national boundaries and facilitate international trade of goods. China's cross-border e-commerce has experienced rapid growth, leveraging e-commerce platforms to achieve cross-border circulation of goods, funds, and information. These platforms face challenges such as product information aggregation, user behavior analysis, intelligent recommendation, payment settlement, and more. Personalized recommendation systems serve as a pivotal strategy to enhance the core competitiveness of cross-border e-commerce platforms [4].

3 Development of Data Mining-Based Recommendation System

3.1 Data Preprocessing

Before proceeding with model training and mining, it is essential to preprocess the raw user behavior data and product data to enhance data quality. Data preprocessing typically involves steps like data cleaning, data integration and merging, and data transformation. Data cleaning detects and eliminates missing values, duplicate data, outliers, and noisy data to reduce errors (refer to Equation 1).

$$\mathrm{Df}_{\mathrm{clean}} = \mathrm{Df.dropna()} \tag{1}$$

Among these, Dfclean represents the cleaned dataset, whereas Df stands for the original dataset.

Data preprocessing involves integrating data from diverse sources like logs, transactions, and user info into a unified dataset. It also transforms raw data into a standardized format suitable for model training, handling time, location, text, etc. Sampling can reduce dataset size for faster computation. The goal of preprocessing is to improve data integrity, accuracy, and consistency, establishing a clean data foundation for subsequent analysis. In summary, data preprocessing aggregates, transforms, and samples data to enable effective data mining and analysis [5]. Refer to Figure 1 for illustration.

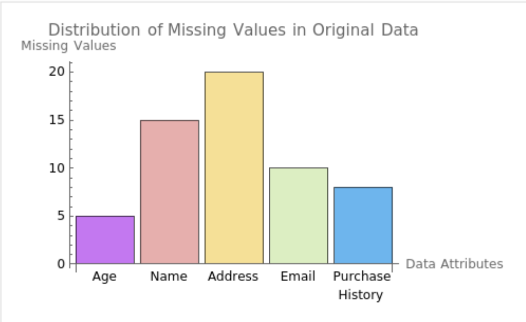

Fig. 1. Distribution of Missing Values in the Original Data

3.2 Feature Engineering

Feature engineering refers to the process of extracting and processing features from data based on business requirements, constructing meaningful feature vectors to provide information input for subsequent model training and prediction (refer to Equation 2).

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{2}$$

Where x' represents the scaled feature value and x is the original feature value.

For structured data like product info and user profiles, features can be directly extracted. For unstructured data like text and images, processing is needed to extract features. Common feature types include product attributes, user demographics and behavior, and user-product interactions. Techniques like feature combination, generalization, PCA dimension reduction, and domain knowledge incorporation can expand and refine features. Effective feature engineering enhances model training performance, improving prediction accuracy [6]. Refer to Figure 2 for visualization.

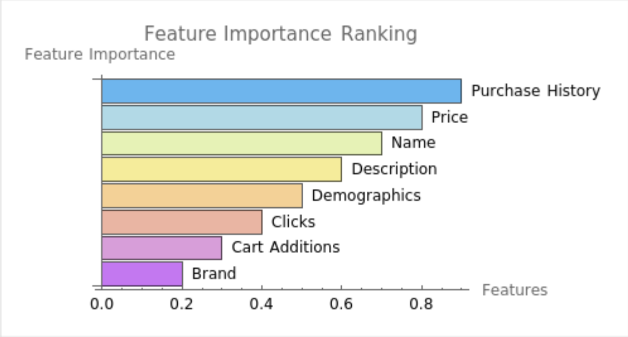

Fig. 2. Ranking of Different Feature Importances

3.3 Recommendation System Models

Depending on the business scenario and dataset characteristics, different recommendation algorithms can be selected and customized to build the system model.

The user-based collaborative filtering algorithm can be represented as (refer to Equation 3):

$$p(u, i) = \sum_{v \in N(u,i)} sim(u, v) \times r(v, i) \tag{3}$$

Where p(u, i) represents the predicted rating of user u for item i, sim(u, v) is the similarity between users u and v, and r(v, i) is the actual rating of user v for item i.

For content-based recommendations, product feature vectors can be used to calculate similarity and recommend similar items. Collaborative filtering establishes user-user or item-item similarity matrices from user behavior data for personalized recommendations. Deep learning models can capture latent user interests from behavior data. Hybrid recommendation systems combine content-based, collaborative filtering, and deep learning models. In practice, careful model tuning and continuous online optimization are needed [7].

4 System Implementation

4.1 Overall Design

The system uses a streaming Spark architecture with modules for data collection, ETL processing, regular model training/optimization, and real-time recommendations. The data source collects user behavior data. ETL processing transforms the data. Model training regularly trains and tunes models on processed data. The recommendation module utilizes the models to provide real-time recommendations for user requests.

4.2 Module Design and Implementation

(1) Data Source Module: Collects user behavior data from website server logs, user login information, transaction records, and more.

```
def collect_data_from_logs(log_file):
    with open(log_file, 'r') as f:
        logs = f.readlines()
    return logs
```

Data Processing Module: Utilizes Spark for data cleaning, integration, and Extract-Transform-Load (ETL) operations. Hive is employed for data warehousing storage.

```
from pyspark.sql import SparkSession
spark = SparkSession.builder.appName("RecommendationSystem").getOrCreate()
def process_data(data):
    df = spark.read.json(data)  # Assume that the data is in JSON format
    # Data cleaning, integration and other operations
    clean_df = df.na.drop()
    return clean_df
```

Model Training Module: Utilizes Spark MLlib for model training, enabling easy implementation of content-based recommendation, collaborative filtering, and other algorithms.

```
from pyspark.ml.recommendation import ALS
def train_model(data):
    als = ALS(userCol="userId", itemCol="itemId", ratingCol="rating")
    model = als.fit(data)
    return model
```

Online Recommendation Module: Leveraging the trained model, this module constructs indexes and labels user requests, thereby delivering personalized recommendation services.

```
def recommend_for_user(model, user_id, num_recommendations):
    recommendations = model.recommendForAllUsers(num_recommendations).filter(col('userId') == user_id)
    return recommendations
```

The system adopts mainstream big data processing frameworks, employs modular design, and each module has well-defined responsibilities. This design enables the effective realization of the recommendation system, encompassing the entire process from data to model and then to service [8].

5 Testing and Evaluation

5.1 Test Case Design

Comprehensive testing requires sufficient test cases covering normal, boundary, invalid inputs and diverse usage scenarios. Test cases should validate functionality, performance, stability and other metrics. They should represent different user types and simulate real-world actions like browsing, searching, clicking, purchasing to cover major interaction scenarios. In summary, thorough test case design is crucial to test diverse scenarios and validate all aspects of the recommendation system.As shown in Table 1.

Table 1. Different types of test cases

Test type	Scene description	Input/operation	Expected result
Normal stream input	Browse products, search, buy	Different user operations	Display correct information, order generation
Boundary value input	Product quantity boundary, search keywords	Boundary value, empty string	The corresponding information or error message is displayed
Abnormal input	Incorrect payment, authentication failure	Invalid payment methods and vouchers	Display error message

5.2 Evaluation Metrics

Key evaluation metrics for recommendation systems include precision, recall (accuracy & coverage), hit rate (process effectiveness), coverage, diversity (result set breadth). Novelty, user satisfaction are also useful. By comparing metric results against business objectives and thresholds, system performance can be quantitatively assessed.

Comprehensive metrics reflecting accuracy, coverage, novelty, diversity, and user satisfaction are crucial for evaluating recommendation systems [9].

5.3 Analysis of Test Results

After conducting comprehensive testing, it's essential to gather statistical data for each metric and analyze the system's performance against the set threshold requirements. If the results do not meet the standards, adjustments to recommendation algorithms and parameters are necessary for optimization. Simultaneously, investigating the reasons behind the issues and proposing targeted solutions is crucial. This optimization process involves iterative steps of testing, result analysis, parameter adjustments, redeployment, and so on, carried out continuously until all evaluation metrics reach the established business objectives [10].

6 Conclusion

In summary, this paper outlines methods for developing intelligent recommendation systems to enhance cross-border e-commerce competitiveness. Data mining techniques like classification, clustering, and association rules enable personalized recommendations. A systematic development approach is needed, including data preprocessing, feature engineering, model training, and online recommendation modules. Comprehensive testing, evaluation, and continuous optimization are crucial. The paper demonstrates the feasibility of leveraging data mining to build intelligent recommendation systems for cross-border e-commerce. Key insights on system design and development are provided. In conclusion, utilizing data mining techniques to develop intelligent recommendation systems is an effective strategy to enhance cross-border e-commerce competitiveness.

References

1. Da'U A, Salim N, Idris R. Multi-level attentive deep user-item representation learning for recommendation system[J]. Neurocomputing, 2021, 433:119-130.
2. Wang T Z. The Ontology Recommendation System in E-Commerce Based on Data Mining and Web Mining Technology[J].Springer Berlin Heidelberg, 2012.
3. Peng G. Recommendation system research on e-commerce based on data mining technology[J]. Heilongjiang Science, 2016.
4. Li R. Construction of E-Commerce Intelligent Recommendation System based on User Behavior Mining[J]. 2016.
5. Sun,Li. The Application Design of Personalized Recommendation System Based on Data Mining to E-Commerce[J].Advanced Materials Research, 2014, 989-994:4538-4541.
6. Deng J, Qing H. Cross-border E-commerce Course Construction Based on Data Mining Algorithm[J]. Journal of Physics: Conference Series, 2021, 1852(2):022040 (7pp).
7. Sustainability. Online Recommendation Systems: Factors Influencing Use in E- Commerce [J]. 2020.

8. Carneiro A. Using Web Data for Measuring the Effectiveness of an E-Commerce Site[J]. 2022..
9. Zhang X, Xu D, Xiao L. Intelligent Perception System of Big Data Decision in Cross-Border e-Commerce Based on Data Fusion[J]. Hindawi Limited, 2021.
10. Song R J, Liang Y, Zhang X H. The Research of E-Commerce Recommendation System Based on Collaborative Filtering[M].IEEE Computer Society,2012.