A Personalized Recommendation Method Based on GAN+XGBoost+LR

Authors: Da Yupeng, Chen Yanchun, Chen Yanchun

Date: 2023-05-16T00:00:00+00:00

Abstract

Objective: To address the problem of insufficient sample data for long-tail item recommendation, where existing collaborative filtering methods suffer from high computational cost and inadequate performance.

Method: We propose a GAN+XGBoost+LR-based solution that employs collaborative filtering to obtain initial sample data, utilizes GAN to generate additional sample data for training the XGBoost+LR model, and devises targeted training strategies tailored to the characteristics of different models.

Results: This solution enhances the robustness of recommendation models while balancing performance and accuracy requirements.

Limitations: The XGBoost model has limited capability for automated feature engineering.

Conclusion: The GAN+XGBoost+LR-based personalized recommendation method can improve the robustness of long-tail item recommendation.

Full Text

Personalized Recommendation Method Based on GAN+XGBoost+LR

Da Yupeng[1],[2], **Chen Yanchun**[1]
[1] School of Economics and Management, Shijiazhuang Tiedao University, Shijiazhuang 050043
[2] Information Center, Hebei Provincial Department of Human Resources and Social Security, Shijiazhuang 050071

Abstract:
[**Objective**] To address the challenges of limited sample data for long-tail item recommendations, high computational costs of existing collaborative filtering methods, and inadequate performance.

[**Methods**] We propose a GAN+XGBoost+LR solution that uses collaborative filtering to identify initial sample data, leverages GAN to generate additional training samples for the XGBoost+LR model, and employs tailored training strategies for different model components.
[**Results**] This approach enhances recommendation model robustness while meeting both performance and accuracy requirements.
[**Limitations**] The XGBoost model's capacity for automated feature engineering is limited.
[**Conclusions**] The GAN+XGBoost+LR personalized recommendation method improves robustness in long-tail item recommendations.

Keywords: machine learning; personalized recommendation; Generative Adversarial Networks; XGBoost; long-tail goods
Classification Number: TP393

Introduction

The exponential growth of e-commerce products makes it increasingly difficult for users to locate desired items. Information overload forces users to expend considerable effort selecting from massive amounts of data [1]. Recommendation systems offer a common solution to this problem by learning from users' historical behavioral preferences to actively predict their potential reactions to unrated or unreviewed items, automatically and personally presenting recommendation lists to help users find needed or preferred products more efficiently. Consequently, achieving cold start for long-tail items under resource constraints and high efficiency requirements has become a critical research focus [2].

Throughout the evolution of personalized recommendation methods, item cold start has remained a significant challenge. Approaches can be categorized into machine learning-based and deep learning-based methods. Machine learning methods further divide into similarity-based and model-based techniques. Similarity-based methods predict based on similarities between users or items—for example, books by the same author or on the same topic, or male users from the same region. Model-based approaches construct predictive models from user click or purchase behavior data. Deep learning methods leverage powerful feature extraction and nonlinear fitting capabilities for recommendations [3]. The cold start problem encompasses two aspects: new item cold start and new user cold start. For new users, the absence of historical rating, review, or behavioral data renders many recommendation methods that rely on user history ineffective (user cold start). For new items, the lack of reviews, views, transactions, clicks, or ratings makes it difficult to recommend these items to interested users, resulting in item cold start.

Numerous scholars have proposed solutions to this problem. Zahid et al. [4] introduced a normalization technique for item details in neighborhood models to model user engagement and preferences for cold start scenarios, using normal-

ization to rank new items and users by relevant tag information and map them to a common latent feature space for baseline rating normalization, thereby addressing cold start with limited data. Tahmasebi et al. [5] proposed a hybrid recommendation method based on profile expansion techniques to alleviate new user cold start, creating a denser user-item rating matrix by adding supplementary ratings and providing target users with enhanced rating profiles. Paleti et al. [6] developed a community detection-based alternating least squares factorization method that combines Louvain algorithm and alternating least squares advantages, using Louvain to analyze user relationships and alternating least squares for prediction. Yadav et al. [7] presented a recommendation system leveraging linked open data and social network features, constructing user profiles based on collaborative features from Linked Open Data (LOD) and social network characteristics to solve pure new user cold start.

Overall, existing solutions fall into two categories: The first approach mines correlations between products or users, linking new products with existing ones and new users with existing users to leverage existing recommendation models. The second approach enriches data for new products or users using third-party information such as social network data and product tags, then constructs predictive models. This paper addresses the cold start problem for long-tail goods by proposing a personalized recommendation method based on GAN+XGBoost+LR.

Note: Figure translations are in progress. See original paper for figures.