METHODS FOR TACKLING COLD START PROBLEM IN RECOMMENDER SYSTEMS

Ryerson University

Supervisor: Ceni Babaoglu

Course: Big Data Analytics

Code: CIND 820

Prepared By: Abdur Rahman Mahbub

Student ID: 501142642

Submitted On: 28th September, 2023

Table of Contents

Abstract (Updated)

The recommender systems help users discover preferred items in modern e-commerce platform, making it an essential part of the digital economy. It is a significant machine learning model that learns from historical data and auxiliary source to provide personalized user recommendations, increasing user engagement and resulting in higher satisfaction. It plays a pivotal role in almost all kinds of e commerce platform However, due to some systematic constraints, these systems often face difficulties in delivering reliable results over all hampering its primary objective. Users need to search for preferred items from a wide range of items and products and the system needs to execute that preference from different subcategories to satisfy that need. When it comes to a highly dynamic platform, many users typically join and adding new items makes it difficult for the system to predict appropriate outcomes. Cold Start Problem directly impacts the recommender system to perform accurately and deliberately force it to be unreliable.

Introduction

Despite the advantages of recommender system, recommender systems often grapple with the cold start problem. This issue arises when the system struggles to provide accurate recommendations for new users or items with limited historical data, primarily due to a lack of user interaction data. This problem typically arises in the following two context.

1. **User Cold Start:** When a new user interacts or appears on the platform without any history or context, the system fails to understand the preferences of the user due to lack of information and suggest irrelevant items.

2. **Item Cold Start:** When a new item is introduced into the vast product catalog, there is a lack of historical user data associated with this item. The system confuses about how to categorize or response to that particular item as they don't have any prior history.

this project proposal aims to propose a pertinent technique to tackle cold start problem in recommender systems, a challenge that holds significant implications for customer satisfaction and business success in the e-commerce platform. By employing the desired dataset and a blend of techniques and tools, this project aspires to offer practical solutions to improve the precision and effectiveness of recommendations, catering to both new and existing users.

Problem Statement

Optimization of product recommendations is a vital issue we seek to address here. To solve problems like handling new product and user introductions, improving the accuracy of recommendations, this project aims to propose a pertinent technique to tackle cold start problem in recommender systems which can give users more personalized and up-to-date recommendations.

Research Questions

The following research questions have been identified to guide this project:

1. How demographic and behavioral data can be leveraged to provide user specific recommendations mitigating cold start problem?

2. How can we use novel strategies when the user interaction is minimal to none or new items are entering the market to enhance the accuracy and personalized nature of product or item recommendations?

3. How can we assess the performance of the advanced recommender systems comprehensively and ensure that the provided recommendations are accurate for a diverse user base?

Dataset Selection

For this project, we have primarily selected the UC Irvine Machine Learning Repository's Online Retail Data Set. After reviewing and analyzing rigorously, MovieLens dataset seemed to be more aligned with the project preference and its applicability for mitigating cold start problem. Public access to the dataset is available and can be viewed or accessed using the following link:

https://grouplens.org/datasets/movielens/

There are several dataset available with the title “MovieLens”. For this project, "MovieLens Latest Small" dataset is selected due to its relevance, richness of data, availability, scalability, and applicability. It is the most suitable dataset, as it demonstrates real scenarios that give an idea of how possible mitigation techniques could be adopted to address the cold start problem.

Techniques

The following techniques and instruments will be used to respond to the research issues, as well as develop more sophisticated recommendation systems for retailers' products.

- **Accommodating data and Preprocessing**

  Collect and preprocess transactional data from the predefined dataset, cleaning and transforming it into a format suitable for recommendation modelling.

- **Collaborative Filtering**

  Implementing collaborative filtering methods based on items and individuals to classify customer preferences and generate product recommendations.

- **Matrix Factorization**

  Exploring matrix factorization techniques to capture itemized interactions, latent factors.

- **Hybrid Recommender Systems**

  Combining the capacity of both content-based and collaborative filtering to provide more accurate recommendations, especially for users with limited interaction history.

- **Deep Learning**

  Experimenting recommendation models to leverage sequential user behaviours and improve personalization.

- **Evaluation Metrics**

  Using common evaluation metrics to measure the effectiveness of the recommender system.

Tools

In the areas of data preparation, model development and evaluation, we will use industry-standard tools and libraries to efficiently deal with cold start problems in recommender system. For data analysis, modelling and evaluation, Python will be the main programming language. Data manipulation and machine learning, as well as collaborative filtering will be facilitated through libraries and frameworks such as pandas, Scikitlearn or Surprise. The efficient handling of large quantities of data will benefit from cloud computing resources.

Reference

F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems (TiiS) 5, 4: 19:1–19:19. https://doi.org/10.1145/2827872