جامعة جدة
University of Jeddah

CCAI422 Recommender System Course Project

Phase 1
Paper and Dataset selection

Supervisor Dr.Safa Alsefri, Dr.Latifa Al-jiffry
by Bedoor Ayad, Raneem Alomari, Deema Al-saygh

جامعة جدة
University of Jeddah

CCAI422 Recommender System Course Project

Phase 1: Paper and Dataset selection

PAPER SELECTION

TITLE:
Variational Autoencoders for Collaborative Filtering.

ABSTRACT:
The paper proposes the use of Variational Autoencoders (VAEs) for collaborative filtering in recommender systems. It extends VAEs to handle implicit feedback data and introduces a multinomial likelihood function that is well-suited for modeling click data. The paper also adjusts the standard VAE objective to improve performance. The proposed approach outperforms state-of-the-art baselines on real-world datasets. This non-linear probabilistic model enables us to go beyond the limited modeling capacity of linear factor models which still largely dominate collaborative filtering research.
Variational autoencoders (VAEs) are applied to collaborative filtering tasks. VAEs aim to learn a low-dimensional representation of the input data while simultaneously generating new samples from this learned representation. In the context of collaborative filtering, VAEs can be used to learn the latent factors that capture the preferences of users and the characteristics of items.

JOURNAL:
The paper was published 16 Feb 2018 . You can access the article on the journal's website [1] .

جامعة جدة
University of Jeddah

CCAI422 Recommender System Course Project

Phase 1: Paper and Dataset selection

PAPER DATASET

MOVIELENS DATASET:
All ratings are contained in the file (ratings.csv). Each line of this file after the header row represents one rating of one movie by one user, and has the following format: userId,movieId,rating,timestamp. The lines within this file are ordered first by userId, then, within user, by movieId. Ratings are made on a 5-star scale, with half-star increments (0.5 stars - 5.0 stars). Timestamps represent seconds since midnight Coordinated Universal Time (UTC) of January 1, 1970.

DATASET LINK:
MOVIELENS 9000 MOVIES DATASET.[2]

DATASET JUSTIFICATION:
The MovieLens dataset is a popular benchmark dataset widely used in the field of recommendation systems and collaborative filtering research. It contains movie ratings provided by users, along with other relevant information such as movie titles, genres, and user demographics. The dataset is often split into training and testing sets, allowing researchers to train and evaluate their recommendation models.

To evaluate the performance of the implemented VAE technique, the MovieLens dataset can be used as a suitable testing ground. Researchers can train a VAE model on a subset of the dataset, using user-item interactions (such as ratings) as input. The trained VAE can then generate recommendations for users based on their learned representations. The recommendations can be compared against the ground truth ratings in the testing set to measure the accuracy and effectiveness of the VAE-based collaborative filtering approach.

CCAI422 Recommender System Course Project

Phase 1: Paper and Dataset selection

NEW DATASET SELECTION

NETFLIX DATASET:

The Netflix dataset consists of two main files: the Movie file and the Rating file.

1. Movie File: Contains Movie_ID, Name, Year
   - Movie_ID: A unique identifier assigned to each movie in the dataset.
   - Name: The title or name of the movie.
   - Year: The year in which the movie was released.

2. Rating File: Contains Movie_ID, User_ID, Rating
   - Movie_ID: The identifier of the movie that the rating corresponds to.
   - User_ID: The unique identifier assigned to each user in the dataset.
   - Rating: The rating given by the user for the movie, ranging from 1 to 5 stars.

DATASET LINK:

Netflix-movie-rating-dataset

DATASET JUSTIFICATION:

The Netflix dataset is a popular benchmark dataset widely used in the field of recommender systems and collaborative filtering research. It contains movie ratings and other information. The Netflix dataset is a valuable resource for researchers working on recommendation models. Its large-scale, real-world nature, rich metadata, benchmarking potential, and focus on personalization make it a comprehensive and diverse dataset for training and evaluating recommendation algorithms. The dataset is often divided into training and testing sets, allowing researchers to train and evaluate their recommendation models. By leveraging this dataset, researchers can advance the state-of-the-art in recommendation systems, leading to more accurate and effective recommendations for users in various domains.

CCAI422 RECOMMENDER SYSTEM COURSE PROJECT

PHASE 1: PAPER AND DATASET SELECTION

TEAM CONTRIBUTIONS:

Name	ID	Work Tasks
Bedoor Ayad	2005961	Report writing
Raneem Alomari	2006352	Paper Selection & Used Data Documentation
Deema Al-sayegh	2006085	Data Selection