Explanation

The study focuses on developing and applying a recommendation system for books using different approaches such as content-based filtering, collaborative filtering and neural collaborative filtering (NCF). To begin the study, the study understands the structure and distribution of the dataset, which is 10,000 books with titles, authors, ISBNs, ratings breakdown and years the books were published. Key trends regarding the distribution of average ratings, most rated books, top authors, dominant languages, and temporal patterns in publication years were highlighted using visualizations. This helped in providing a first look at the dataset, and helped in selecting the features for the recommendation models. The first model implemented is a content-based filtering system that used TF-IDF vectorization on combined textual features (book titles and author names). Then, cosine similarity is used to calculate closeness between books to generate the recommendation based on the semantic textual similarity. Even with such an easy model to look at, the model initially had trouble distinguishing between user input variations, and normalizing titles to ensure they match consistently.

Second, the collaborative filtering based on user interaction data is examined using a user-item matrix made from a different rating dataset. In this approach, books similar to the ones rated positively by a user were recommended based on cosine similarity between items (i.e., books). For the second, a user-based recommendation system is also developed based on computing user similarity and predicting a potential rating if a particular user is similar to another in the preferences of similar users. With this method, the customized recommendations are more, giving more than the content-based method, but it is not without data sparsity and cold start got new user and new items.

To test the effectiveness of the collaborative filtering method, the train and test set split is conducted by each user contributing one known rating to the test set. The recommender system achieved modest performance using precision, recall and F1-score metrics and Precision@10 has 0.0306 and Recall@10 has 0.2205. The results for these challenges were the ability to generate highly relevant recommendations from sparse and implicit user feedback.

To improve the recommendation quality, the study implemented a Neural Collaborative Filtering model using TensorFlow and Keras. The way the study uses this model is by encoding users and items into dense embeddings, and then processing through dense neural layers to predict user ratings. It is trained from encoded user-book pairs and ratings associated with them. Preprocessing is applied to users and books, embedding layers were trained to obtain latent features for users and books, and finally the two were concatenated and forwarded through fully connected layers. Firstly, mean squared error and Mean absolute error were used to evaluate the model on a test split and it achieved progressively better performance across epochs with final mean validation MAE of 0.6.

The refinement improved robustness of the model. In terms of flexibility and adaptability, the NCF model renders the traditional approaches old fashioned, and would make a great candidate for real world recommender systems that have nonlinear interactions between users and items. The study concludes with a comprehensive review of different recommendation techniques applied to book data. It provides a methodological progression starting from basic statistical exploration to deep learning-based modeling towards improving the prediction accuracy and personalization. The interpretability of content-based filtering, the personalization of collaborative filtering, as well as the scalability and flexibility of the neural model were each advantages to each method. Through this type of layered approach, the recommender system can reflect diverse user behavior and data environment in the real world.