CMPE-256 Project proposal

Team 1 (Hotel coders): Long Haitao, Aparna Ramesh, Devarsh Thaker

Traveller's guide: A Personalized Hotel Recommender System

This project will create a personalized hotel recommendation system using the HotelRec dataset, which has data from TripAdvisor reviews [1]. The system will suggest hotels based on users' preferences, past experiences, and specific hotel features. We will use both collaborative filtering and content-based methods to give high-quality recommendations.

The HotelRec dataset has 50 million reviews from TripAdvisor. It includes the hotel URL, user information, review date, overall rating, review title, review text, and sub-ratings for specific hotel features like sleep quality, value, rooms, service, cleanliness, and location. This dataset allows us to explore different ways to recommend hotels based on ratings, the attributes of the hotels, and the sentiments expressed in the reviews.

The recommendation system will implement SVD and SVD++ as baseline models. Beyond these, we plan to use Factorization Machines (FM) as algorithms. FMs are well suited for this dataset as they can model interactions among various features like the user, hotel attributes, and sub-ratings [2]. Next, we will use Neural Collaborative Filtering (NCF), which utilizes neural networks to capture nonlinear interactions between users and hotels. Finally, we will use a hybrid approach combining collaborative and content-based filtering by including hotel attributes. This hybrid model will deliver personalized recommendations that reflect both ratings and feedback.

The evaluation of our hotel recommendation system will focus on predictive accuracy and relevance. We will use Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for models. This approach will allow us to evaluate how well our predictions align with user ratings. Moreover, we will implement hyperparameter tuning through grid search to fine-tune model performance and improve the accuracy of our recommendations. Additionally, we will try to evaluate precision and recall at various recommendation depths (such as Precision@K and Recall@K), which will help measure the ranking quality and the system's ability to recommend relevant hotels within the top recommendations.

In conclusion, this project aims to develop a hybrid hotel recommendation system that combines collaborative filtering, content-based filtering, and cutting-edge neural network techniques. The system will offer personalized hotel recommendations tailored to user preferences and the unique characteristics of different hotels. Moreover, the analysis will feature a comparison between the baseline and advanced models, showcasing the improvements in both performance and personalization.

References:

1. D. Antognini and B. Faltings, "HotelRec: A novel very large-scale hotel recommendation dataset," *Proc. of The 12th Language Resources and Evaluation Conference (LREC),* Marseille, France, 2020, pp. 4917–4923, [Online], https://aclanthology.org/2020.lrec-1.605.pdf
2. S. Rendle, "Factorization Machines," *IEEE International Conference on Data Mining*, pp. 1329-1000, 2010, [Online], https://doi.org/10.1109/ICDM.2010.127