Hotel Recommendation System Using Content-Based Filtering

Group 2: Qian Li, Xin Ding, Wenyu Yang, Zhengyi Xu

Project Overview

This project aims to develop an intelligent hotel recommendation system for Seattle visitors using natural language processing (NLP) and content-based filtering techniques. The system will analyse hotel descriptions and characteristics to provide recommendations based on textual similarity, helping travellers find accommodations that best match their preferences without requiring historical user data, thus solving the cold-start problem common in traditional recommendation systems.

Problem Statement

Current hotel recommendation systems face three critical challenges: the cold-start problem for new users, time sensitivity of historical data not reflecting current conditions, and information overload making manual comparison time-consuming. Our solution addresses these limitations by providing real-time, contextual recommendations based on concrete hotel attributes and characteristics, ensuring suggestions align with users' present interests rather than potentially outdated historical behaviour.

Data Source

The project utilizes a comprehensive dataset of Seattle hotels compiled from major booking platforms (Booking.com, Expedia, and TripAdvisor) containing detailed property descriptions, amenities, and location information. This data is structured in a CSV file (Seattle_Hotels.csv) that serves as our primary data source, ensuring a rich foundation for content-based analysis.

AI Methods

The system implements three main technical components:
(1) Text Preprocessing: including stop-word removal, text cleaning, normalization, and N-gram generation to capture meaningful phrases
(2) Feature Extraction: utilizing TF-IDF vectorization and potential word embeddings to capture semantic relationships
(3) Similarity Computation: employing Cosine Similarity metrics and weighted similarity measures to generate accurate recommendations based on content similarity.

Evaluation Methods

The system's performance will be assessed through both quantitative and qualitative measures: Quantitative evaluation using precision metrics and verification against pre-labelled lists of similar hotels; Qualitative analysis through manual comparison of recommended properties and direct user feedback on recommendation relevance. This dual approach ensures both systematic accuracy and practical usefulness of the recommendations while allowing for continuous refinement based on user expectations.

Implementation Considerations

Project success depends heavily on data quality, computational resources for real-time processing, and the effectiveness of our NLP pipeline. The system must generate recommendations promptly for a seamless user experience, while maintaining flexibility to incorporate user feedback for ongoing improvements. Regular updates to the hotel database will ensure the continued relevance and accuracy of recommendations.