

CSC111 Project Proposal: Restaurant Recommendation System

Qingyi Jiang, Songxuan Wu, Rachel Yeung, Jiaqi Zhao

Monday, March 3, 2025

Problem Description and Research Question

TODO

Computational Plan

Use of Trees and Graphs

Decision Tree for Filtering

- A decision tree is used to filter restaurants based on user preferences, such as cuisine type, budget range, and minimum rating.
- The tree structure follows a hierarchical filtering mechanism:
 - **Root Node:** Represents all available restaurants.
 - **First Level:** Categorization based on cuisine type.
 - **Second Level:** Further categorization based on budget.
 - **Third Level:** Final filtering based on minimum rating.
 - **Leaf Nodes:** Restaurants that meet all criteria.
- This approach allows efficient filtering before applying ranking algorithms.

Graph-Based Ranking with PageRank

- A graph is constructed where:
 - **Nodes** represent individual restaurants.
 - **Edges** connect restaurants with similar attributes (cuisine, ratings, and pricing).
 - **Edge Weights** are computed based on similarity scores derived from restaurant attributes.
- We apply PageRank to determine the relative importance of each restaurant in the network, ensuring that more "popular" restaurants (based on customer reviews and connectivity) are ranked higher.

Minimum Spanning Tree (MST) Optimization (Optional)

- As a refinement step, we use an MST-based approach to ensure that recommended restaurants maintain a degree of coherence.
- MST computation follows these steps:
 1. Construct a graph where nodes are restaurants, and edges are weighted by similarity.
 2. Compute the Minimum Spanning Tree (MST) to cluster related restaurants.
 3. Adjust PageRank results to reflect MST connectivity.
- This step ensures that recommendations are cohesive and logically structured, rather than isolated results.

Dataset and Sample Data

Our project uses a real-world restaurant dataset containing the following key attributes:

- **Name:** Restaurant name.
- **Address:** Physical location of the restaurant.
- **Online Order & Booking Availability:** Whether the restaurant offers online orders and table booking.
- **Rating:** Customer rating (out of 5) based on reviews.
- **Restaurant Type:** Categorization such as casual dining, fine dining, buffet, etc.
- **Cuisines:** Types of cuisine served (e.g., North Indian, Chinese, Mughlai, Thai, etc.).
- **Approximate Cost:** Estimated cost for two people.
- **Dish Liked:** Popular dishes among customers.

Name	Cuisines	Rating	Cost	Restaurant Type
Jalsa	North Indian, Mughlai	4.1	800	Casual Dining
Spice Elephant	Chinese, North Indian, Thai	4.1	800	Casual Dining

Types of Computation

1. Data Preprocessing

- Handle missing values and standardize categorical fields.
- Convert cuisine types into numerical encodings where needed.
- Normalize numerical attributes (ratings, votes, pricing).

2. Decision Tree Filtering

- Filter restaurants based on Cuisine Type \rightarrow Budget \rightarrow Minimum Rating.

3. Graph-Based Ranking with PageRank

- Construct a graph representation of the filtered restaurants.
- Compute restaurant similarity using cuisine, ratings, and cost.
- Apply PageRank to rank restaurants based on their connectivity in the similarity graph.

4. (Optional) MST-Based Refinement

- Construct an MST of the restaurant similarity graph.
- Adjust the PageRank ranking to prioritize more connected restaurants.

Use of Python Libraries

pandas (Data Processing)

- **Usage:** Load, clean, and manipulate restaurant data.
- **Key Functions:** `pandas.read_csv()`, `pandas.DataFrame.query()`, `pandas.DataFrame.sort_values()`

networkx (Graph Construction & Ranking Computation)

- **Usage:** Construct a graph representation of restaurants.
- **Key Functions:** `networkx.Graph()`, `networkx.pagerank()`, `networkx.minimum_spanning_tree()`

plotly (Visualization)

- **Usage:** Display restaurant rankings using bar charts.
- **Key Functions:** `plotly.express.bar()`, `plotly.graph_objects.Figure.update_layout()`

Expected Output and Visualization

- Basic Table Display: Restaurants ranked by PageRank score.
- Ranking Visualization: Bar charts using Plotly.
- Interactive Filtering (Optional): Users adjust preferences dynamically via Streamlit.
- Graph Visualization (Optional): Interactive graph representation of restaurant connections.

References

TODO