Introduction

A Hybrid Model is developed for hotel recommendation. The benchmark approach in the industry relies on solely hotel popularity. This method does not provide personalized hotel recommendation. The Hybrid Model combines user preferences and hotel location specific popularity. Recommendation on returning user considers their historical booking history. While for new user, a decision tree is applied to classify them into our existing user clusters.

Method

Utility Matrix:

Giving for each user-item pair, a value that represents what is known about the degree of preference of that user for that item.

Hierarchical Clustering

Begin with every points in its own cluster. Larger clusters will be constructed by combining two smaller clusters.

SVD (decomposition):

SVD seeks a low-rank matrix \( X = U \Sigma V^T \) to reduce the dimension and fill in the some unrated items.

Decision Tree Classifier:

We applied the method of decision tree to classify new users into our user clusters according to their user profiles.

Result & Conclusion

Hybrid model results in prediction with 53.6% accuracy on testing data - 4% improvement on content-base model. This result is consistent with our hypothesis: both user preference and hotel popularity are vital in recommendation system.

Currently, the leading team on Kaggle hotel recommendation competition achieved 51.7% accuracy. The benchmark content-based model has 49.8% accuracy.

Therefore, this Hybrid Model is very likely to get the first place in this competition.

User Profile Detection

Country and region – CORRELATE with users preference.

City - LITTLE effects on users preference.

GUI Design

Accuracy Comparison

<table>
<thead>
<tr>
<th></th>
<th>Hybrid Model</th>
<th>Content-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>61.29%</td>
<td>51.03%</td>
</tr>
<tr>
<td>Test</td>
<td>53.62%</td>
<td>51.74%</td>
</tr>
</tbody>
</table>

We compared the outcomes of predicted top 5 hotels clusters that a coming user might book between single content-based model and hybrid model. Combined with utility matrix in collaborative filtering, the hybrid model reached more accuracy net matter in training or testing to predict the booking outcomes of users.

Reference


