RelaTree

Graph Based approach for Recommendations

Team - RelaTreevs

Aashish Subramanian • Chintan Vachhani • Pavana Srinivasadeshika Achar

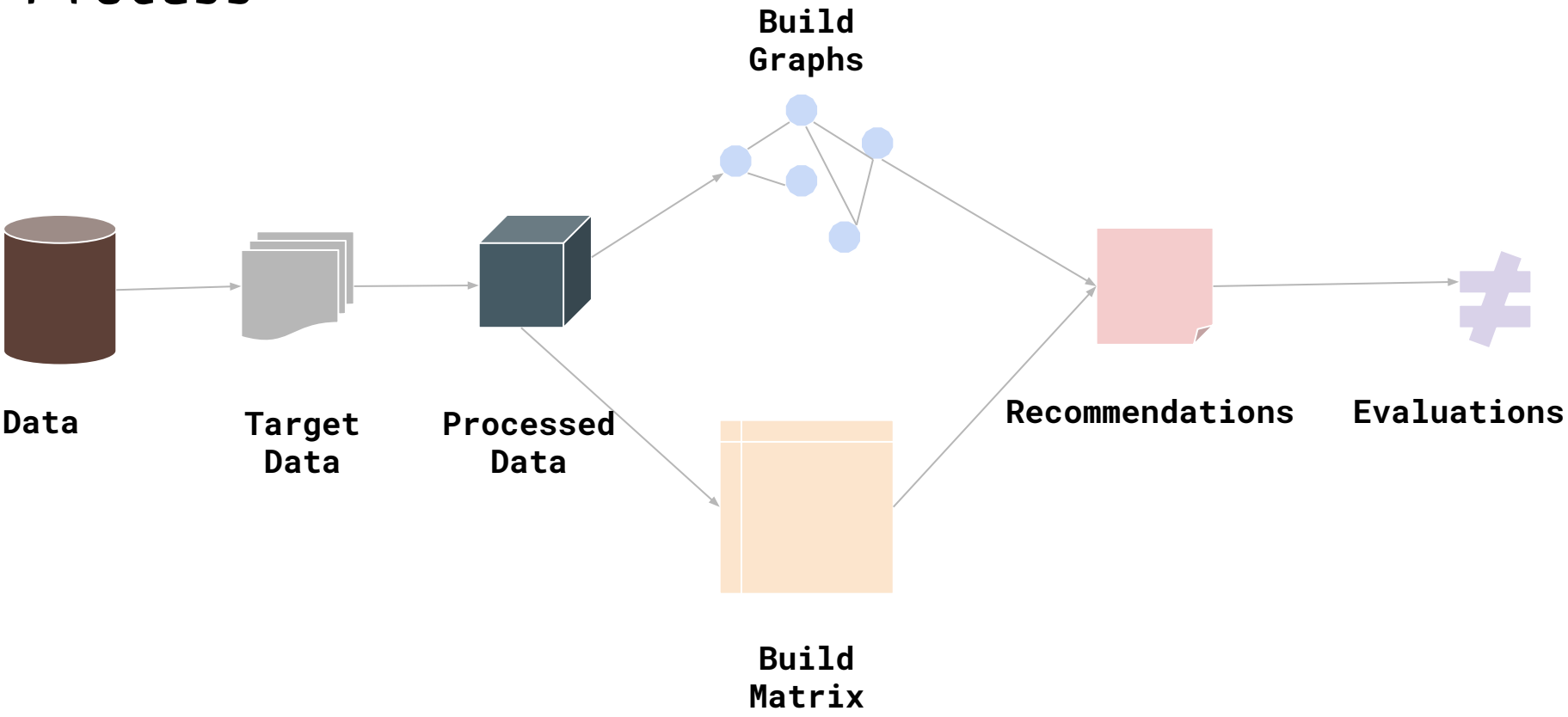
Introduction

- **Problem:** provide personalized recommendations to groups for various services.
- **Goal:** Use relationships between users, groups and services.
- **Solution:** Graph based approach to provide recommendations based on interests of a users' immediate connections.
- **Evaluation:** Comparative analysis performed with item-based recommendation.

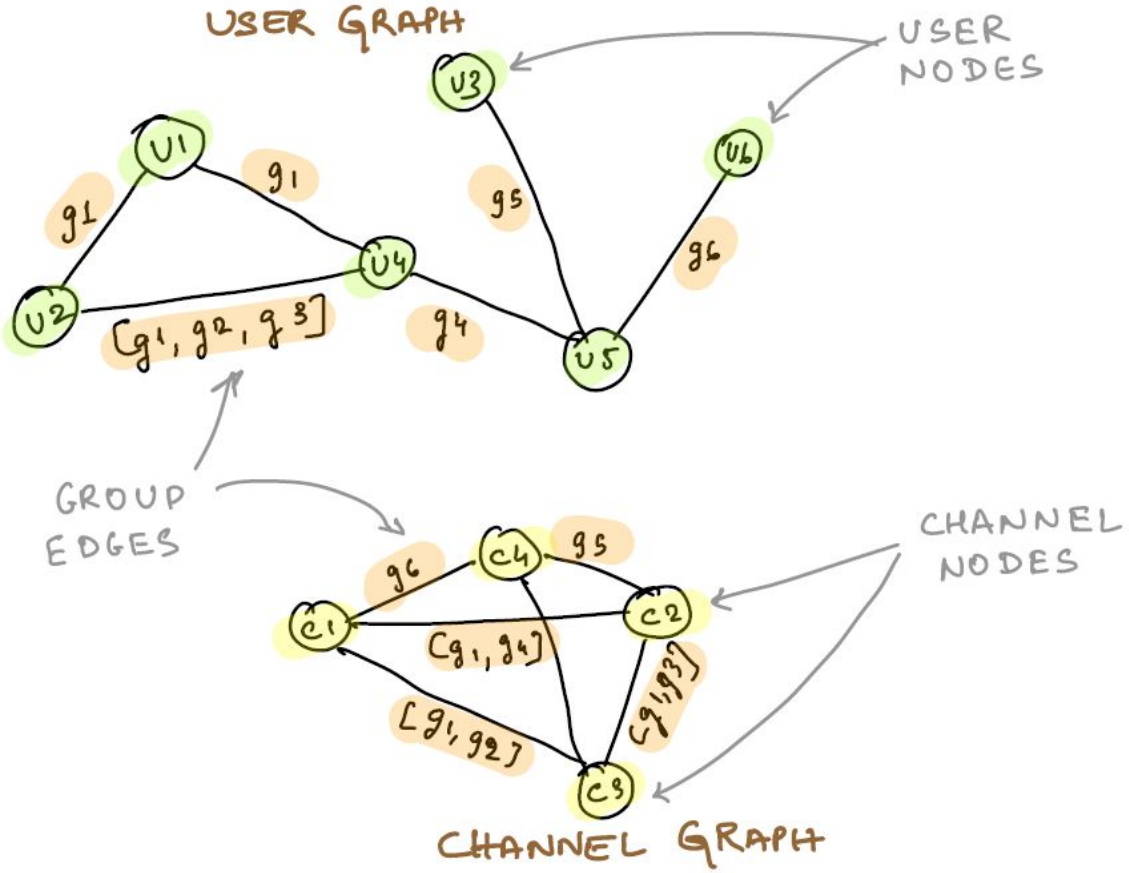
Dataset

- Obtained from multiple tables in the production database of Duta Inc.
- Data has been anonymized.
- Consists of user, group and channel information.
- Each user is connected to a group.
- Each group is subscribed to 1 or more channels.
- The user to group relation consists of 20917620 records.
- The group to channels relation consists of 13150793 records

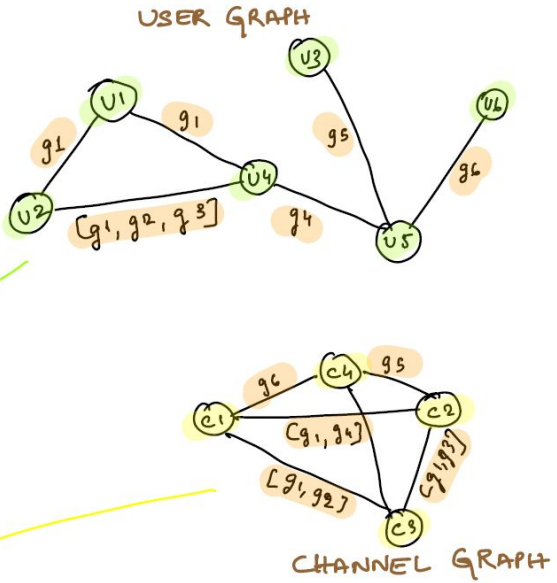
Process


Build
Graphs
Data
Target
Data
Processed
Data
Recommendations
Evaluations
Build
Matrix

Model

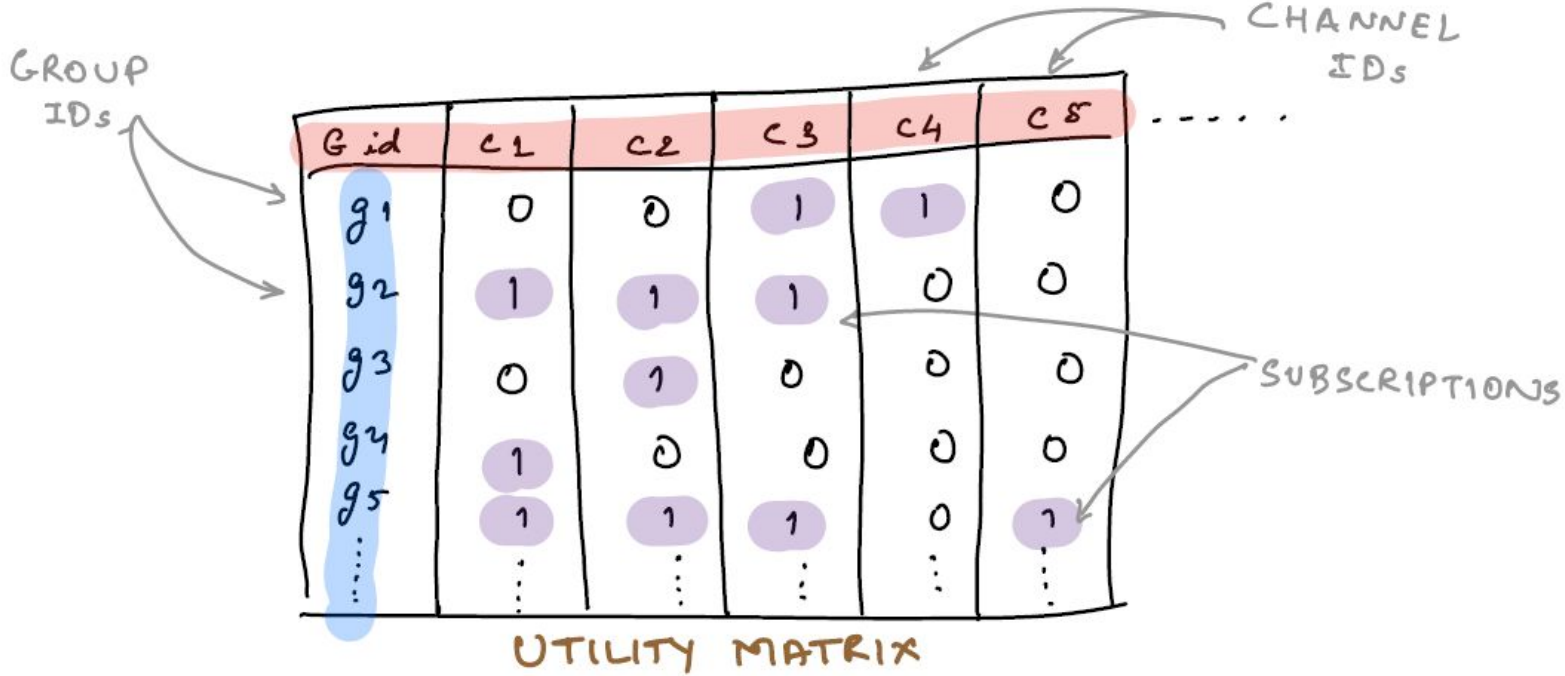

USER GRAPH
USER NODES
U1
U2
U3
U4
U5
U6
g1
g1
g5
g6
g4
[g1, g2, g3]
GROUP EDGES
CHANNEL NODES
C1
C2
C3
C4
g6
g5
[g1, g4]
[g1, g2]
[g1, g3]
CHANNEL GRAPH

Approach

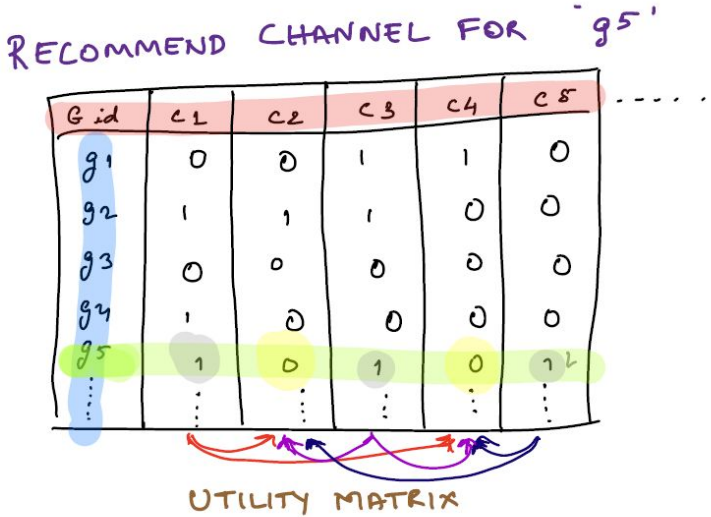


- Get all users for the given group
- Find their 1st connects & create a set of users
- Find a set of groups for all the users along with a weight associated with them based on user counts.
- Find a set of channels for all the groups.
- Recommend 'k' channels using the group weights and channel association scores.

Model


GROUP IDs
CHANNEL IDs
G id
c1
c2
c3
c4
c5
g1
g2
g3
g4
g5
SUBSCRIPTIONS
UTILITY MATRIX

Approach



$$Sim[C_2] = Sim(C_2, C_1) + Sim(C_2, C_3) + Sim(C_2, C_5) + \dots$$

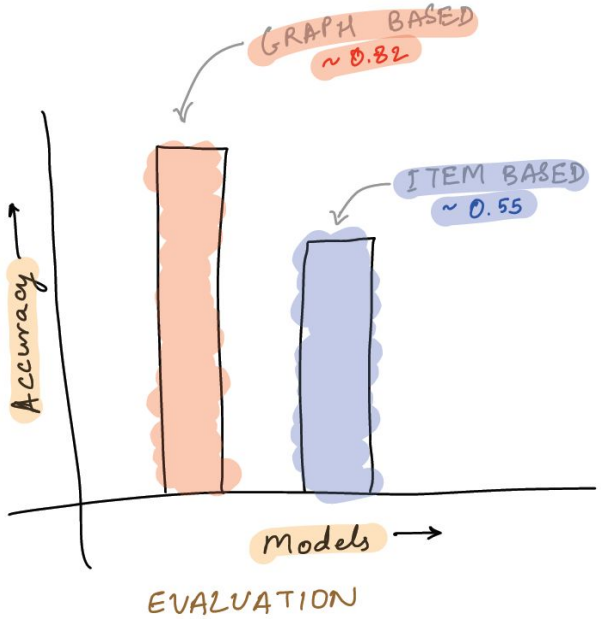
$$Sim[C_4] = Sim(C_4, C_1) + Sim(C_4, C_3) + Sim(C_4, C_5) + \dots$$

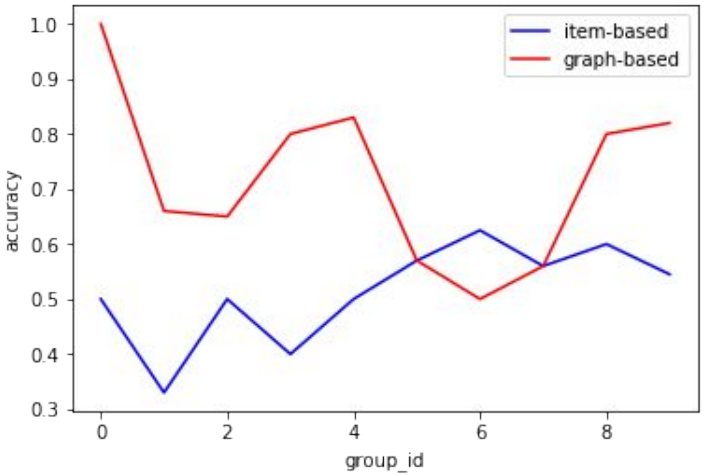
RECOMMEND MAX (Sim[C2], Sim[C4])

Evaluation

- Both the models were evaluated using 10 diverse groups, out of which 3 were weakly connected groups, 5 were heavily connected groups and 2 were lone groups.
- Reason for choosing a very small number of groups for end evaluation is because of the large data. It takes approximately 45 minutes to run for a single weakly connected group.
- But the used set of test sample represent the entire data statistics well. Also, the actual evaluation would be how the live users react to these recommendations.
- Accuracy was used as the metric since subscription to a service is a binary value and Jaccard similarity was used.

Evaluation


GRAPH BASED
~ 0.82
ITEM BASED
~ 0.55
Accuracy
Models
EVALUATION


item-based
graph-based
accuracy
group_id
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
2
4
6
8

Conclusion

- It was observed that graph based recommendation works better overall. It outperforms the traditional method for weakly connected and lone groups.
- Using graph based recommendation, the cold start problem is addressed and it no longer affects the system heavily.
- Scope for improvement exists. Considering additional features such as user’s activity with a particular service, more user activity data and analyzing user patterns can add to the effectiveness of recommendations.

Thank You