

Assignment 2 Part 1

Friend Recommendation using Mutual Friends

Algorithm description:

1. Identify Possible Friends: For every user in the dataset, search for possible friends by looking at the friends of their existing friends. Exclude the user's current friends and the user themselves from this list.
2. Focus on Target Users: Narrow down the process to the 10 target users for whom friend recommendations are needed.
3. Count Shared Friends: For each target user, calculate the number of mutual friends they have with their potential friends.
4. Combine Potential Friend Data: Gather the mutual friend counts for each target user and their possible friends into a unified list.
5. Rank and Choose Top 10: Sort the potential friends for each target user by the mutual friend count in descending order and select the top 10.
6. Present Recommendations: Display the final results in an organized manner.

Detailed Algorithm Description:

Input:

data_rdd: RDD with each entry as "UserID\tlist_of_friends".

userID_list: List of 10 target user IDs.

data_dict: Broadcast variable mapping user IDs to their friends.

Steps:

1. Load and Process Data: Load the friend recommendation file into data_rdd, split each entry into (UserID, list_of_friends), and handle empty friend lists.
2. Generate Potential Friends: Use a flatMap to create pairs of (UserID, PotentialFriendID) where potential friends are friends-of-friends, excluding existing friends and the user.
3. Filter by Target Users: Filter out the pairs where UserID is in userID_list.
4. Count Mutual Friends: Use reduceByKey to count the mutual friends between each UserID and PotentialFriendID.
5. Aggregate Recommendations: Group potential friends by UserID, and aggregate the counts of mutual friends into lists.
6. Sort and Select Top 10: Sort potential friends by mutual friend count and select the top 10 for each user.
7. Format and Display Results: Convert the recommendations into a structured format with UserID, top 10 friend recommendations, and mutual friend counts. Sort and display the results.

Output:

UserID	Friend Recommendations
7	12343, 16539, 40423, 1, 10, 11, 12, 10287, 10469, 12344
58	72, 1, 10, 11, 12, 10782, 10841, 11192, 11599, 12663
569	4839, 7651, 13411, 579, 7611, 10459, 10469, 12715, 22199, 10407
1001	439, 1100, 3895, 442, 13829, 18163, 10240, 11045, 13835, 20598
5891	5883, 46410, 46438, 21100, 21859, 46341, 46377, 46380, 46386, 46390
7498	1250, 5144, 7280, 7296, 7458, 7459, 7460, 7461, 7462, 7463
24167	9774, 24140, 10176, 10333, 10500, 10531, 10532, 10723, 10801, 11066
32491	15175, 16862, 16878, 17740, 23550, 25199, 34185, 35605, 35630, 35692
42165	42152, 42154, 42160, 42167, 42170, 42171, 42172, 42179, 10661, 14327
48651	11234, 31463, 11162, 11245, 9106, 10753, 28354, 44454, 10413, 11697