

CprE 419 Lab 10: Data-Driven Advertising

Shuo Wang

Algorithm:

Rank the users based on the compensation price and how many acquaintances he/she have.

The lower his compensation price is and the more compensation he has, the more valuable the user is.

So rank the users by $\text{rate} = \text{compensation price} / \text{number of acquaintances}$.

The lower the rate is, the higher the rank is.

Then we keep choosing the high rank user until the budget runs out.

(Note: This is a simple algorithm. It does not count for the rate change of the unselected users after one user has been selected. But I hope it can provide a reasonably good result.)

MapReduce:

Two rounds:

Round1: For each user, list all the consumers he could reach and count the total number of these consumers. Then calculate the rate of each user.

Round2. Sort each record based on the user's rate in ascending order. Then keep picking the users with high rank until the budget runs out. Put all the recipients into a HashSet and output the size.

The output is part-r-00000

Line schema: user: ID price: Compensation people: list of acquaintances

The final statistic is:

total number of consumers:50085, total number of users:777, total budget used: 9994.51543042033