Mining massive Datasets WS 2017/18

Problem Set 8

Rudolf Chrispens, Marvin Klaus, Daniela Schacherer

December 13, 2017

Exercise 01

a) The probability that C1 and C2 are matched?

QuastionnairesCount = b (Q_1, ... , Q_b)

QuestionCount = r

QuestionCountSum = $b * r$

MatchCondition = if(mQ_x(r) == $fQ_x(r)$) //m = person1, f = person2, x = element of b

$Probability - of - an - event - happening = \frac{Number\ of\ ways\ it\ can\ happen}{Total\ number\ of\ outcomes}$

Propability that 1 Questionair matsches: $(\frac{1}{r})^r * b$

HIER BRAINSTORME ICH EINFACH NICHT BEACHTEN :D

b) The probability that exactly two (no matter which) questionnaires match, i.e. have the same answers for both C1 and C2.

Exercise 02

MISSING

Exercise 03

MISSING

Exercise 04

- BALANCE Algorithm:
  For each query, assign it to an advertiser with the largest unspent budget (i.e. largest BALANCE) A(x,y) B(x,z)
  Worst Case Szenarios:

a) xyyy - AA__ (since there are 3 y no optimal solution possible, A has only 2 dollars.)

b) xyyx - A_AB (optimal solution is possible but not certain BAAB)

c) yyxx - AABB (Optimal)

d) xzyz - BBA_ (optimal solution is possible but not certain ABAB)

- Since we used the worst case possible query assignment with the balance Algorithm we can show that only c) would give an optimal solution. All the other queries have situations where an optimal solution is not certain. c's optimal solution is because A and B cant steal from each other, because their budget is empty before the other can take anything from their common x.

Exercise 05

- budget: 3 / ties in favor of lower index

- $A_1(Q_1,Q_2)$ $A_2(Q_2,Q_3)$ $A_3(Q_3,Q_4)$ $A_4(Q_1,Q_4)$

- Query: $Q_1, Q_2, Q_3, Q_3, Q_1, Q_2, Q_3, Q_1, Q_4, Q_1, Q_4$

a) What is the sequence of advertisers that the BALANCE algorithm will yield?
$A_1, A_2, A_3, A_2, A_4, A_1, A_3, A_4, A_3, A_1, A_4$
What is the competitive ratio for this instance?
$Competitive\ ratio = min_{(}all\ possible\ inputs)\frac{|M_{(greedy)}|}{(|M_{(opt)}|)})$

b) Rearrange the sequence of queries so that BALANCE results in a worse competitive ratio.
Query: $Q_1, Q_1, Q_1, Q_1, Q_4, Q_4, Q_2, Q_2, Q_3, Q_3, Q_3$
$A_1, A_4, A_1, A_4, A_3, A_3, A_2, A_2, A_2, A_3,$ __

Exercise 06

MISSING

Exercise 07

MISSING