Tyler Steele and Robert Albright
May 2, 2007

Recommendation Algorithms

Utility – sum over all users of probability that users will buy recommended item.

$$\Pi = \sum_u P_i(u)$$

$A = PW$

$\Pi(OPT) = \Sigma$ max element in each row of A

Assume we consider only the first two items of each purchase. Each transaction corresponds to an edge between two items.

Special case:

1) only two classes of items c_1 and c_2
2) preference for c_1 summed over all users ≥ preference c_2 summed over all users

Three Algorithms

VIC – Vote in Cluster:
  If purchase is two items for some category, recommend that category
  If purchase is from both categories, recommend c_1

VOC – Vote Out of Cluster:
  Always vote for c_1

VRC – Vote Random Cluster:
  If purchase is two items for some category, recommend that category
  If purchase is from both categories, recommend either randomly with equal prob.

Example where VOC outperforms VIC

Two types of users, two clusters, one item in each cluster.

$$A = PW = \begin{pmatrix} .9 & .1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Assume 10^3 as many type 1 users as type 2 users.

Only need to calculate utility for both items in c_2 since utility is same in other cases

VIC – (#users type 1)(prob. type 1 users selects cluster 2)(value of recommend.) + (#users type 2)(prob. type 2 users selects cluster 2)(value of recommend.)

$= 1000*10^{-2}*10^{-1} + 1*1*1 = 2$

VOC $= 1000*10^{-2}*.9 + 1*1*0 = 9 > 2$

ALG = some algorithm

OPT = optimal solution

Measure of algorithm would be $\min_{all\ data} \frac{\Pi(ALG)}{\Pi(OPT)}$ but that value changes based on input.

Claim: No algorithm can do better than $\frac{2}{\sqrt{k}+1}$ where k is #clusters.

Claim: VRC achieves this bound.

Theorem: If sample size s = 2, then

1) $\forall$ alg no matter how good$|\ \min_{all\ data} \frac{\Pi(\text{alg})}{\Pi(\text{opt})} \leq \frac{2}{\sqrt{k}+1}$

2) $\min_{all\ data} \frac{\Pi(\text{VRC})}{\Pi(\text{opt})} = \frac{2}{\sqrt{k}+1}$

Proof: $\Pi(\text{opt}) = \sum_u \max_i A_{ui}$ $\qquad$ $P_i(u) = A_{ui} = P_i$ when u is implied

For user u, both elements are in c_i with probability P_i^2 and utility P_i

A cross edge has prob. $P_i * P_j$ and utility $\frac{P_i + P_j}{2}$

$$\Pi(\text{VRC}) = \sum_u \left[\sum_{i=1}^{k} P_i^3 + \sum_{i \neq j} P_i P_j \cdot \frac{P_i + P_j}{2} \right]$$

¿ $$\sum_u \left[\sum_{i=1}^{k} P_i^3 + \frac{1}{2} \left(\sum_{i \neq j} P_i^2 P_j + \sum_{i \neq j} P_i P_j^2 \right) \right]$$

¿ $$\sum_u \left[\sum_{i=1}^{k} P_i^3 + \sum_{i \neq j} P_i P_j^2 \right]$$

¿ $$\sum_u \left[\sum_{i=1}^{k} P_i^3 + \sum_{i} P_i^2 \sum_{j \neq i} P_j \right]$$

¿ $$\sum_u \left[\sum_{i=1}^{k} P_i^3 + \sum_{i} P_i^2 (1 - P_i) \right]$$

¿ $$\sum_u \left[\sum_{i=1}^{k} P_i^3 + \sum_{i} P_i^2 - \sum_{i} P_i^3 \right]$$

¿ $$\sum_u \left[\sum_{i} P_i^2 \right]$$

so …

$$\min_{\text{all data}} \frac{\Pi(\text{VRC})}{\Pi(\text{OPT})} = \min_{\text{all data}} \frac{\sum_u \sum_i P_i{}^2}{\sum_u \max_i P_i} = \min_{\text{all data}} \frac{\sum_u \left(\left(\frac{1-P_{\max}}{k-1} \right)^2 (k-1) + P_{\max}{}^2 \right)}{\sum_u \max_i P_i}$$

$\frac{f(b)}{b}$ minimizes $\frac{a_i}{b_i}$ where $a_i = f(b_i)$.

Out minimum occurs at $\frac{1}{\sqrt{k}}$, and value is $\frac{2}{\sqrt{k}+1}$