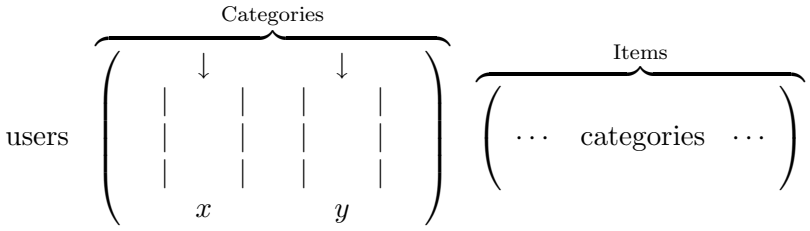
CS 485 - Lecture 38

Saudia Ooyshee

May 2, 2006

A store might tell you what you want to buy and we have this model $A = PW$.



Nice thing is, we could get data for the matrix W from all users assuming we knew what the categories were. Now question is how do we determine categories? We don't want to deal with the various number of items a customer buys. So, we are going to go thorough the cash register and if a customer buys only 1 item, we are going to throw that receipt away. If a customer buys 2 or more items, we are going to randomly pick 2 items and say that constitues the purchase. So, we are going to assume every customer always buys 2 items.

Sample size $s = 2$; This is number of items purchased in a transaction.
$\#(i)$ = number of times i purchased
$\#(i, j)$ = number of times pair(i, j) purchased
Do i and j belong in same category?

$$\text{freq of } i = \sum_u P_{uc(i)} W_{c(i)i} = W_{c(i)i} \sum_u P_{uc(i)}$$

$$\text{freq of pair}(i,j) = \sum_u P_{uc(i)} W_{c(i)i} P_{uc(j)} W_{c(j)j}$$

Note: Difference in order you do the summation.
We want to consider two vectors - x and y.

$$x = (P_{1c(i)}, P_{2c(i)}, \cdots, P_{mc(i)})$$

$$y = (P_{1c(j)}, P_{2c(j)}, \cdots, P_{mc(j)})$$

If i and j are in the same category then $x = y$. If x and y are not close to parallel then categories i and j are distinguishable.

$$\frac{E^2(\#(i,j))}{E(\#(i,i))E(\#(j,j))} = \frac{(W_{c(i)i} W_{c(j)j} \sum_u P_{uc(i)} P_{uc(j)})^2}{[W^2_{c(i)i} \sum_u P_{uc(i)} P_{uc(i)}][W^2_{c(j)j} \sum_u P_{uc(j)} P_{uc(j)}]}$$

$$= \frac{(x \times y)^2}{(x \times x)(y \times y)}$$

$$= \frac{(x \times y)^2}{x^2 y^2}$$

$$= \cos^2 \theta$$

θ is the angle between x and y. If they are same $\cos^2 \theta$ will be 1; if they are not identical then cosine will be something significantly less than 1.

Example

$$PW = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \underbrace{\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}}_{(a \quad b \quad c \quad d \quad e \quad f)}$$

Assume that each of the three buyers visists the store equally often.

$$Prob(a) = \frac{1}{3} \times 1 \times \frac{1}{3} + \frac{1}{3} \times 0 \times \frac{1}{3} + \frac{1}{3} \times \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$$

--	a	b	c	d	e	f
Prob	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

If things were statistically independent, the probability that some pair was chosen is given by the table below:

	a	b	c	d	e	f
a	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{24}$	$\frac{1}{48}$	$\frac{1}{48}$
b						
c						
d						
e						
f						

But if we look at the cash register receipts, we will find out that this is not the frequency of the pair.

For ordered pair: What's the probability of pair(d, e) ?

$$Prob[(d, e)] = \frac{1}{3} \times 0 \times 0 + \frac{1}{3} \times (1 \times \frac{1}{2}) \times (1 \times \frac{1}{4}) + \frac{1}{3} \times (\frac{1}{2} \times \frac{1}{2}) \times (\frac{1}{2} \times \frac{1}{4})$$

$$= \frac{1}{24} + \frac{1}{96} = \frac{5}{96}$$

$$Prob[(a, d)] = \frac{1}{9 \times 8}$$

$$Prob[(d, d)] = \frac{5}{3 \times 16}$$

$$Prob[(e,e)] = \frac{5}{3 \times 64}$$
$$Prob[(a,a)] = \frac{5}{3 \times 36}$$
$$Prob[(d,e)] = \frac{5}{3 \times 32}$$

Are d and e in the same category?

$$\frac{P^2[(d,e)]}{P[(d,d)] \times P[(e,e)]} = \frac{(\frac{5}{3\times 32})^2}{(\frac{5}{3\times 16})(\frac{5}{3\times 64})} = 1$$

So, d and e are in the same category.
Are a and d in the same category?

$$\frac{P^2[(a,d)]}{P[(a,a)] \times P[(d,d)]} = \frac{(\frac{1}{9\times 8})^2}{(\frac{5}{3\times 36})(\frac{5}{3\times 16})} = \frac{1}{25}$$

So, a and d are in different category.

Some Recommendation Algorithms

Continuing to use a sample size of 2, we can build a graph where the vertices represent the items we have and each edge represents a user/transaction where the end points of the edge are the items that the user bought. For now, assume that there are 2 disjoint categories, C_1 and C_2, and all items in a category are equally likely. Without lost of generality, also assume that C_1 is preferred over C_2 (or that they are equally preferred).

Neighbor Algorithm: If a user selects items a and b, look for another user who selected either a or b, and recommend the other item that the second user selected.

Voting Algorithm: Recommend item c such that the number of times users select (a, c) or (b, c) is maximized over all c.

Vote In Cluster (VIC): If a user selects two items in C_1, let's call this a C_1-edge. If the items a user selects are in different categories, call that a cross-edge. For each edge, if it's a C_1-edge or cross-edge, vote for C_1. Otherwise, vote for C_2. Recommend an item from the category that gets the most votes.

Vote Out of Category (VOC): Vote for C_1 no matter what purchase was.

Vote Randomly (VRC): For a C_1 edge, vote for C_1. Otherwise, vote randomly.