

Clustering

$N = \# \text{ of likes} = 482,055$

$U = \# \text{ of unique users} = 49,093$

$P^* = \# \text{ of possible user profiles} = 3,920$
 $= 2 \cdot 5 \cdot 7 \cdot 7 \cdot 8$

$P = \# \text{ of actual user profiles} = 1,520$

$P^* - P = 2,400$

$P/P^* = 38.8\%$

<u>i</u>	<u>Profile</u>	<u>Cat1</u>	<u>Cat2</u>	- - - -	<u>Cat 41</u>
1	00000	$P_{1,1}$	$P_{1,2}$	- - -	$P_{1,41}$
\vdots	\vdots	\vdots			
3920	14667	$P_{3920,1}$	$P_{3920,2}$	- - -	$P_{3920,41}$

$$\forall i \sum_{j=1}^{41} P_{i,j} \approx 1$$

$$\sum_{j=1}^{41} C_{i,j} \approx 1$$

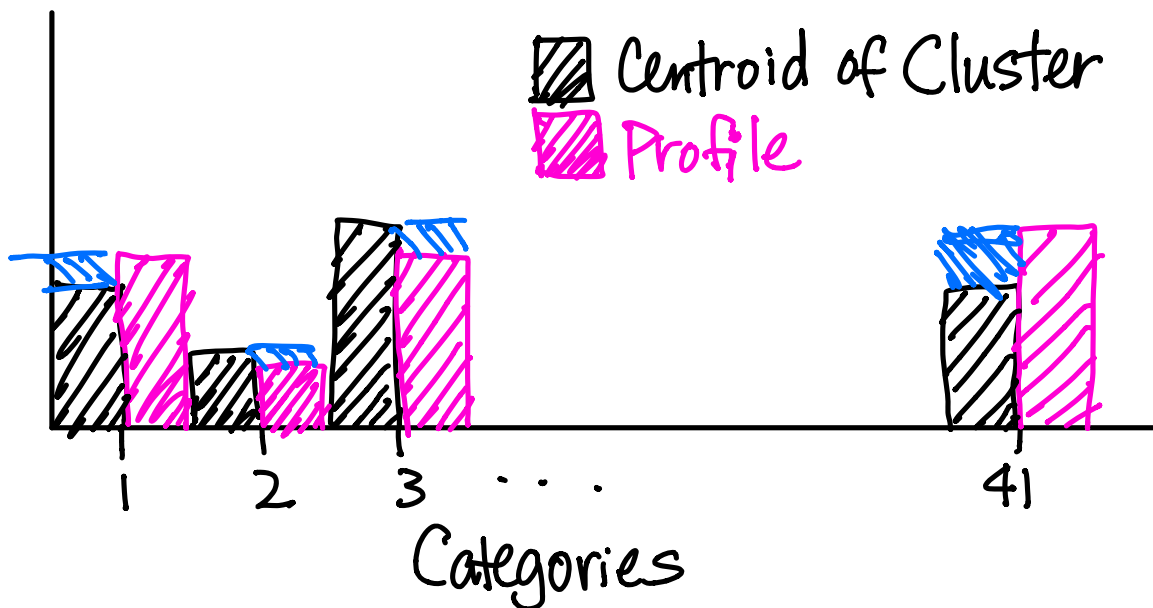
<u>Cluster</u>	<u>Cat1</u>	<u>Cat2</u>	- - -	<u>Cat 41</u>
1	$C_{1,1}$	$C_{1,2}$	- - -	$C_{1,41}$
2	$C_{2,1}$	$C_{2,2}$	- - -	$C_{2,41}$
\vdots	\vdots			
k	$C_{k,1}$	$C_{k,2}$	- - -	$C_{k,41}$

$D_{a,b}$ = distance between Profile a
and Cluster b

$$= \sqrt{\sum_{j=1}^{41} (P_{a,j} - C_{b,j})^2 \cdot C_{b,j}}$$

Profile a belongs to Cluster b

if $D_{a,b} = \min \{D_{a,1}, D_{a,2}, \dots, D_{a,k}\}$



* For profiles with no like data,

assign randomly to cluster containing the most profiles with at least 3 common attributes

e.g. 01234 has no like data ↘

Cluster 1

01200 ✓

01134 ✓

12347 ✗

Cluster 2

00000 ✗

11111 ✗

12222 ✗

Cluster 3

11235 ✓

02236 ✓

01235 ✓

