

演習5

学号：_____

姓名：_____

(演習5-1)

- 次の説明文について、用語一覧から正しい英字を選んで記入してください

- (**s. 推薦システム**)とは、ユーザの好みに合わせて、おすすめのアイテムを提示してくれるシステムである。これは**ユーザの評価・行動履歴**にもとづいて推薦を行う (**q. 協調フィルタリング**) アイテムの内容 (属性) にもとづいて推薦を行う (**b. 内容ベースフィルタリング**) とに大別される。
- 前者の推薦には、**嗜好が類似しているユーザが好むアイテムを推薦する** (**c. ユーザベース**) と **アイテム評価の類似性にもとづいて推薦する** (**r. アイテムベース**) に分けられる
- また入力データは、主に2つに大別される。1つは、利用者が**アイテムを好きか嫌いかを評価した** (**n. 明示**) 的データがある。
- もう1つは利用者の**行動から好き・嫌いを評価した** (**i. 暗黙**) 的データなどが用いられる。これは例えば、(**k. 閲覧**) データなどである。

用語一覧

- a. OIP
- b. 内容ベースフィルタリング
- c. ユーザベース
- d. 特徴
- e. 情報フィルタリング
- f. ISO
- g. アイテム
- h. 集合知
- i. 暗黙
- j. SEO
- k. 閲覧
- l. 詳細
- m. 検索システム
- n. 明示
- o. PageRank
- p. ハイブリッド
- q. 協調フィルタリング
- r. アイテムベース
- s. 推薦システム

(演習5-2) アイテムベース協調フィルタリング

以下の購買データを用いて関連アイテムを作成してください

User	Item1	Item2	Item3	Item4	Item5
U1	●	-	●	●	-
U2	-	●	●	-	-
U3	-	●	-	●	●
U4	●	-	●	●	-
U5	●	-	●	-	-
U6	-	●	●	-	●
	3	3	5	3	2

共購買の人数

Item (Y\X)	I1	I2	I3	I4	I5
I1	-	0	3	2	0
I2	0	-	2	1	2
I3	3	2	-	2	1
I4	2	1	2	-	1
I5	0	2	1	1	-

共購買の割合

Item (Y\X)	I1	I2	I3	I4	I5
I1	-	0/3	<u>3/5</u>	<u>2/3</u>	0/2
I2	0/3	-	<u>2/5</u>	<u>1/3</u>	<u>2/2</u>
I3	3/3	<u>2/3</u>	-	<u>2/3</u>	<u>1/2</u>
I4	2/3	<u>1/3</u>	<u>2/5</u>	-	<u>1/2</u>
I5	0/3	<u>2/3</u>	1/5	<u>1/3</u>	-

<関連アイテム>

X	Y	共購買 割合
I1	I3	3/3
I5	I2	2/2
I1	I4	2/3
I2	I3	2/3
I2	I5	2/3
I4	I1	2/3
I4	I3	2/3
I3	I1	3/5
I5	I3	1/2
I5	I4	1/2
I3	I4	2/5
I3	I2	2/5
I2	I4	1/3
I4	I2	1/3
I4	I5	1/3

割合
0.3以上



(演習5-3) ユーザAの「1.トマト」の予測評価値

User	1.トマト	2.ナス	3.キュウリ	4.キャベツ	5.レタス	6.ジャガイモ	7.ニンジン	8.ピーマン	9.ネギ	10.カボチャ	評価平均値
A	<u>?</u>	?	0	0	1	0	1	0	1	1	
B	0	1	1	1	0	0	1	1	0	0	5/10=0.5
C	0	0	1	0	1	1	1	0	0	0	4/10=0.4
D	1	0	1	0	1	0	0	0	1	1	5/10=0.5
E	1	0	1	1	1	1	0	1	0	1	7/10=0.7

User	UserAとの類似度
B	0.25
C	0.45
D	0.67
E	0.41

$$P_{a,i} = \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_u) \text{Sim}(a,u)}{\sum_{u \in U} |\text{Sim}(a,u)|}$$

$$\begin{aligned} \underline{P_{A,1}} &= \{(\underline{r_{B,1}} - \text{ave}(r_B)) * \text{sim}(u_A, u_B) + (\underline{r_{C,1}} - \text{ave}(r_C)) * \text{sim}(u_A, u_C) \\ &\quad + (\underline{r_{D,1}} - \text{ave}(r_D)) * \text{sim}(u_A, u_D) + (\underline{r_{E,1}} - \text{ave}(r_E)) * \text{sim}(u_A, u_E)\} \\ &\quad / (|\text{sim}(A,B)| + |\text{sim}(A,C)| + |\text{sim}(A,D)| + |\text{sim}(A,E)|) \end{aligned}$$

$$= \{(\underline{0} - 0.5) * 0.25 + (\underline{0} - 0.4) * 0.45 + (\underline{1} - 0.5) * 0.67 + (\underline{1} - 0.7) * 0.41\} / (0.25 + 0.45 + 0.67 + 0.41)$$

$$= ((-0.125) + (-0.18) + (0.335) + (0.123)) / 1.78 = \underline{0.086}$$

(演習5-4)

- 表 1 は14人 (A~N) のユーザの食べ物の好みの数値です (1:嫌い ~ 5:好き)
- 表 1 を用いて、以下のユーザの類似度を計算してください
 - D と C の類似度 ($\text{Sim}(D,C)$) = **0.94**

表 1

User	1. 寿司	2. ピザ (比萨)	3. ラーメン (汤面)	4. ハンバーガー (汉堡包)	5. 麻婆豆腐	6. 餃子 (饺子)	7. カレー (咖喱)	8. パスタ (面食类)	9. パン (面包)	10. ケーキ (水果蛋糕)	平均評価値
A		1	1	2	3	1	4	1	1	5	2
B	1		4	2	3	2	3	4	3	2	2.7
C	1	1		1	3	2	1	2	1	1	1.5
D	1	1	4		3	3	2	1	1	1	1.9
E	2	2	5	3		4	3	3	5	2	3.2
F	1	2	4	3	3		3	3	5	1	2.9
G	3	2	4	4	1	2		3	3	3	2.9
H	1	1	2	2	3	2	3		3	1	2.2
I	5	2	4	1	4	3	1	3		1	2.5
J	1	1	2	1	5	2	1	2	3		1.9
K	3	5	5	5	1	5	1	3		5	3.4
L	1	1	4	1	4	3	2		4	1	2.5
M	3	5	4	2	3	2		3	1	5	2.9
N	1	1	2	1	2		2	3	4	5	2.3

(ヒント) A と B の類似度

User	1 ・ 寿司	2 ・ ピザ (比萨)	3 ・ ラーメン (汤面)	4 ・ ハンバーガー (汉堡包)	5 ・ 麻婆豆腐	6 ・ 餃子 (饺子)	7 ・ カレー (咖喱)	8 ・ パスタ (面食类)	9 ・ パン (面包)	10 ・ ケーキ (水果蛋糕)
A		1	1	2	3	1	4	1	1	5
B	1		4	2	3	2	3	4	3	2

$$\text{Cos}(x, y) = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2} \sqrt{\sum y_i^2}} = \frac{x_1 y_1 + x_2 y_2 \dots + x_n y_n}{\sqrt{x_1^2 + x_2^2 \dots + x_n^2} \sqrt{y_1^2 + y_2^2 \dots + y_n^2}}$$

$$\begin{aligned} \text{Cos}(A, B) &= (1*4 + 2*2 + 3*3 + 1*2 + 4*3 + 1*4 + 1*3 + 5*2) \\ &\quad / (\text{sqrt}(1^2+2^2+3^2+1^2+4^2+1^2+1^2+5^2)*(\text{sqrt}(4^2+2^2+3^2+2^2+3^2+4^2+3^2+2^2))) \\ &= \underline{\underline{0.75}} \end{aligned}$$

(演習5-5)

- (演習5-4) で求めた類似度を用いて、Dさんの「4.ハンバーガー(汉堡包)」の予測評価値 ($P_{D,4}$) を計算してください

User	1. 寿司	2. ピザ (比萨)	3. ラーメン (汤面)	4. ハンバーガー (汉堡包)	5. 麻婆豆腐	6. 餃子 (饺子)	7. カレー (咖喱)	8. パスタ (面食类)	9. パン (面包)	10. ケーキ (水果蛋糕)	平均評価値(\bar{r}_c)
C	1	1		1	3	2	1	2	1	1	1.5
D	1	1	4	?	3	3	2	1	1	1	1.9
G	3	2	4	4	1	2		3	3	3	2.9

$$P_{D,4} = ((r_{C,4} - \text{ave}(r_C)) * \text{sim}(D,C) + (r_{G,4} - \text{ave}(r_G)) * \text{sim}(D,G)) / (|\text{sim}(D,C)| + |\text{sim}(D,G)|)$$

$$= ((1 - 1.5) * 0.94 + (4 - 2.9) * 0.8) / (0.94 + 0.8)$$

$$= \underline{\underline{0.236}}$$

(ヒント) 協調フィルタリング

(2.類似ユーザの重みを用いて、未評価アイテムの予測評価値を計算)

- 予測評価値 (predict-value)

$$P_{a,i} = \frac{\sum_{u \in U} (r_{u,i} - \text{ave}(r_u)) \text{Sim}(a,u)}{\sum_{u \in U} |\text{Sim}(a,u)|}$$

$P_{a,i}$:	ユーザa のアイテムi の予測評価値
$u \in U$:	ユーザ集合
$r_{u,i}$:	あるユーザu のアイテムi に対する評価
$\text{ave}(r_u)$:	ユーザu の評価の平均

<ポイント> 未評価アイテムについて

- 類似する人の重みを高くして、その類似する人の評価を採用する

(演習5-6)

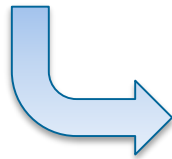
「3.ラーメン（汤面）」が好きな人には、何を推薦すべきですか？

User	1. 寿司	2. ピザ (比萨)	3. ラーメン (汤面)	4. ハンバーガー (汉堡包)	5. 麻婆豆腐	6. 餃子 (饺子)	7. カレー (咖喱)	8. パスタ (面食类)	9. パン (面包)	10. ケーキ (水果蛋糕)
A							●			●
B			●					●		
C										
D			●							
E			●		●	●			●	
F			●			●			●	
G			●	●						
H										
I	●		●		●					
J					●					●
K		●	●	●		●				●
L			●		●				●	
M		●	●							
N			●			●			●	●
人数	1	2	10	2	4	4	1	1	4	4

(演習5-6)

「3.ラーメン (汤面)」が好きな人には、何を推薦すべきですか？

User	1. 味噌 (味噌)	2. ピザ (比萨)	3. ラーメン (汤面)	4. ハンバーガー (汉堡包)	5. 麻婆豆腐	6. 餃子 (饺子)	7. ナカレー (茄腔)	8. キバスタ (韭菜菜)	9. パン (面包)	10. ケーキ (水果蛋糕)
A							●			●
B			●					●		
C										
D			●							
E			●		●	●			●	
F			●			●			●	
G			●	●						
H										
I	●		●		●					
J					●					●
K		●	●	●	●	●				●
L			●		●				●	
M		●	●							
N			●			●			●	●
人数	1	2	10	2	4	4	1	1	4	4



Item	2. ピザ (比萨)	3. ラーメン (汤面)	4. ハンバーガー (汉堡包)	5. 麻婆豆腐	6. 餃子 (饺子)	9. パン (面包)	10. ケーキ (水果蛋糕)
2.ピザ (比萨)	–	2/10	1/2	1/4	0/4	0/4	1/4
3.ラーメン (汤面)	2/2	–	2/2	0/4	4/4	4/4	2/4
4.ハンバーガー (汉堡包)	1/2	2/10	–	0/4	1/4	0/4	1/4
5.麻婆豆腐	0/2	3/10	0/2	–	1/4	2/4	1/4
6.餃子 (饺子)	1/2	4/10	1/2	1/4	–	3/4	2/4
9.パン (面包)	0/2	4/10	0/2	2/4	3/4	–	1/4
10.ケーキ (水果蛋糕)	1/2	2/10	1/2	1/4	2/4	1/4	–

「3.ラーメン (汤面)」が好きな人には、
「6.餃子 (饺子)」「9.パン (面包)」を推薦するのが良い