

ネットワークコンピューティング 中間試験

関西学院大学工学部情報科学科
北村泰彦

1

問題1

	1:吉野屋	2:スキ家	3:なか卯	4:松屋
1:北村	2	1	*	3
2:角所	*	1	3	3
3:長田	3	2	3	1
4:西関	1	1	2	*

上の表は、 $R=\{1,2,3\}$ とする評価値行列である。推薦利用者を1:北村としたとき、以下の問いに答えよ。

- (1) Pearson相関を用いて、北村と角所、北村と長田、北村と西関との相関をそれぞれ示せ。
- (2) 協調フィルタリングの手法を利用して、北村の「なか卯」に対する評価値を推定せよ。

2

問題1: 解答例

(1) Pearson相関を用いて、北村と角所、北村と長田、北村と西関との相関をそれぞれ示せ。

- 1:北村と2:角所の相関を計算する。この二人がともに評価しているアイテムは2:スキ家、4:松屋なので、 $Y_{1,2} = \{2,4\}$ となる。これらのアイテムについての $Y_{1,2}$ 上の平均評価値はそれぞれ以下の通りである。

$$\bar{s}'_1 = \frac{\sum_{k=2,4} s_{1,k}}{2} = \frac{1+3}{2} = 2$$
$$\bar{s}'_2 = \frac{\sum_{k=2,4} s_{2,k}}{2} = \frac{1+3}{2} = 2$$

3

問題1: 解答例

- したがって相関は

$$\rho_{1,2} = \frac{\sum_{k=2,4} (s_{1,k} - \bar{s}'_1)(s_{2,k} - \bar{s}'_2)}{\sqrt{\sum_{k=2,4} (s_{1,k} - \bar{s}'_1)^2} \sqrt{\sum_{k=2,4} (s_{2,k} - \bar{s}'_2)^2}}$$
$$= \frac{(1-2)(1-2) + (3-2)(3-2)}{\sqrt{(1-2)^2 + (3-2)^2} \sqrt{(1-2)^2 + (3-2)^2}}$$
$$= 1$$

- 1:北村と3:長田の相関を計算する。この二人がともに評価しているアイテムは1:吉野屋、2:スキ家、4:松屋なので、 $Y_{1,3} = \{1,2,4\}$ となる。これらのアイテムについての $Y_{1,3}$ 上の平均評価値はそれぞれ以下の通りである。

$$\bar{s}'_1 = \frac{\sum_{k=1,2,4} s_{1,k}}{3} = \frac{2+1+3}{3} = 2$$
$$\bar{s}'_3 = \frac{\sum_{k=1,2,4} s_{3,k}}{3} = \frac{3+2+1}{3} = 2$$

4

問題1: 解答例

- したがって相関は

$$\begin{aligned}\rho_{1,3} &= \frac{\sum_{k=1,2,4}(s_{1,k} - \bar{s}'_1)(s_{3,k} - \bar{s}'_3)}{\sqrt{\sum_{k=1,2,4}(s_{1,k} - \bar{s}'_1)^2} \sqrt{\sum_{k=1,2,4}(s_{3,k} - \bar{s}'_3)^2}} \\ &= \frac{(2-2)(3-2) + (1-2)(2-2) + (3-2)(1-2)}{\sqrt{(2-2)^2 + (1-2)^2 + (3-2)^2} \sqrt{(3-2)^2 + (2-2)^2 + (1-2)^2}} \\ &= -\frac{1}{2}\end{aligned}$$

- 1:北村と4:西関の相関を計算する. この二人がともに評価しているアイテムは1:吉野屋, 2:スキ家なので, $Y_{1,4} = \{1,2\}$ となる. これらのアイテムについての $Y_{1,2}$ 上の平均評価値はそれぞれ以下の通りである.

$$\begin{aligned}\bar{s}'_1 &= \frac{\sum_{k=1,2} s_{1,k}}{2} = \frac{2+1}{2} = \frac{3}{2} \\ \bar{s}'_4 &= \frac{\sum_{k=1,2} s_{4,k}}{2} = \frac{1+1}{2} = 1\end{aligned}$$

5

問題1: 解答例

- したがって相関は

$$\begin{aligned}\rho_{1,4} &= \frac{\sum_{k=1,2}(s_{1,k} - \bar{s}'_1)(s_{4,k} - \bar{s}'_4)}{\sqrt{\sum_{k=1,2}(s_{1,k} - \bar{s}'_1)^2} \sqrt{\sum_{k=1,2}(s_{4,k} - \bar{s}'_4)^2}} \\ &= 0\end{aligned}$$

6

問題1: 解答例

(2)協調フィルタリングの手法を利用して, 北村の「なか卯」に対する評価値を推定せよ.

- $\bar{s}_1 = \frac{\sum_{k=1,2,4} s_{1,k}}{3} = \frac{2+1+3}{3} = 2$

- したがって,

$$\begin{aligned}\hat{s}_{1,3} &= \bar{s}_1 + \frac{\sum_{i=2,3,4} \rho_{1,i}(s_{i,3} - \bar{s}'_i)}{\sum_{i=2,3,4} |\rho_{1,i}|} \\ &= 2 + \frac{(1)(3-2) + (-\frac{1}{2})(3-2) + 0(2-1)}{|1| + |-\frac{1}{2}| + |0|} \\ &= \frac{7}{3}\end{aligned}$$

7

問題2

	属性a	属性b
x_1	6	4
x_2	4	2
x_3	2	0
x_4	2	2
x_5	6	0

K-means法を用いて上記の5つのデータ x_1, \dots, x_5 を三つのクラスターに分割したい. 初期データ分割を $C_1 = \{x_1, x_2, x_3\}, C_2 = \{x_4\}, C_3 = \{x_5\}$ としたときの, 最終分割結果を示せ.

8

問題2: 解答例

	属性a	属性b
x_1	6	4
x_2	4	2
x_3	2	0
x_4	2	2
x_5	6	0

クラスタ数を3, 初期データ分割を
 $C_1=\{x_1, x_2, x_3\}, C_2=\{x_4\}, C_3=\{x_5\}$ とする。

C_1, C_2, C_3 の重心はそれぞれ
 (4,2)
 (2,2)
 (6,0)

距離の2乗

	C_1	C_2	C_3
x_1	8	20	16
x_2	0	4	8
x_3	8	4	16
x_4	4	0	20
x_5	8	20	0

x_3 は C_2 へ移動させる。

9

問題2: 解答例

	属性a	属性b
x_1	6	4
x_2	4	2
x_3	2	0
x_4	2	2
x_5	6	0

クラスタ数を3, データ分割を
 $C_1=\{x_1, x_2\}, C_2=\{x_3, x_4\}, C_3=\{x_5\}$ とする。

C_1, C_2, C_3 の重心はそれぞれ
 (5,3)
 (2,1)
 (6,0)

距離の2乗

	C_1	C_2	C_3
x_1	2	25	16
x_2	2	5	8
x_3	18	1	16
x_4	10	1	20
x_5	10	15	0

終了

10

問題3

「気象条件とハイキングに関するデータ」が与えられたとき, カバーリングアルゴリズムを用いて, ハイキングをするときのルールを示せ.

11

問題3: 気象条件とハイキングに関するデータ(○はハイキングをする, ×はハイキングをしないことを表す.)

天気	温度	湿度	風	ハイキング
晴れ	暑い	高い	無し	×
晴れ	暑い	高い	有り	×
曇り	暑い	高い	無し	×
雨	暖かい	高い	無し	×
雨	涼しい	普通	無し	×
雨	涼しい	普通	有り	×
曇り	涼しい	普通	有り	○
晴れ	暖かい	高い	無し	○
晴れ	涼しい	普通	無し	○
雨	暖かい	普通	無し	×
晴れ	暖かい	普通	有り	○
曇り	暖かい	普通	有り	○
曇り	暑い	普通	無し	○
雨	暖かい	高い	有り	×

12

問題3: 解答例

- 以下のルールを学習することを目標とする。
If [?] Then ハイキング=○
- 未知の項?には, 10の選択肢が存在する.
 - 天気=晴れ: 3/5
 - 天気=曇り: 3/4
 - 天気=雨: 0/5
 - 温度=暑い: 1/4
 - 温度=暖かい: 3/6
 - 温度=涼しい: 2/4
 - 湿度=高い: 2/7
 - 湿度=普通: 5/7
 - 風=有り: 3/6
 - 風=無し: 3/8

13

問題3: 解答例

- 最大正答率の選択肢は「天気=曇り」なので, 以下のルールを選択する。
If [天気=曇り] Then ハイキング=○
- このルールがカバーする範囲は3/4であるので, 新しい条件を付加して, ルールを洗練する。
If [天気=曇り] and [?] Then ハイキング=○

14

天気	温度	湿度	風	ハイキング
曇り	暑い	高い	無し	×
曇り	涼しい	普通	有り	○
曇り	暖かい	普通	有り	○
曇り	暑い	普通	無し	○

15

問題3: 解答例

- ここで未知の項?で考えられる条件項は, 以下の7つ.
 - 温度=暑い: 1/2
 - 温度=暖かい: 1/1
 - 温度=涼しい: 1/1
 - 湿度=高い: 0/1
 - 湿度=普通: 3/3
 - 風=有り: 2/2
 - 風=無し: 1/2
- 「湿度=普通」が最大正答率なので, 以下のルールを得る。
If [天気=曇り] and [湿度=普通] Then ハイキング=○
- このルールは6事例のうちの3事例しかカバーしていないので, 3事例をデータセットから削除し, ルール学習を繰り返す。

16

天気	温度	湿度	風	ハイキング
晴れ	暑い	高い	無し	×
晴れ	暑い	高い	有り	×
曇り	暑い	高い	無し	×
雨	暖かい	高い	無し	×
雨	涼しい	普通	無し	×
雨	涼しい	普通	有り	×
晴れ	暖かい	高い	無し	○
晴れ	涼しい	普通	無し	○
雨	暖かい	普通	無し	×
晴れ	暖かい	普通	有り	○
雨	暖かい	高い	有り	×

17

問題3: 解答例

- 以下のルールを学習することを目標とする。
If [?] Then ハイキング=○
- 未知の項?には, 10の選択肢が存在する.
 - 天気=晴れ: 3/5
 - 天気=曇り: 0/1
 - 天気=雨: 0/5
 - 温度=暑い: 0/3
 - 温度=暖かい: 2/5
 - 温度=涼しい: 1/3
 - 湿度=高い: 1/6
 - 湿度=普通: 2/5
 - 風=有り: 1/4
 - 風=無し: 2/7

18

問題3: 解答例

- 最大正答率の選択肢は「天気=晴れ」なので, 以下のルールを選択する。
If [天気=晴れ] Thenハイキング=○
- このルールがカバーする範囲は3/5であるので, 新しい条件を付加して, ルールを洗練する。
If [天気=晴れ] and [?] Thenハイキング=○

19

天気	温度	湿度	風	ハイキング
晴れ	暑い	高い	無し	×
晴れ	暑い	高い	有り	×
晴れ	暖かい	高い	無し	○
晴れ	涼しい	普通	無し	○
晴れ	暖かい	普通	有り	○

20

問題3: 解答例

- 未知の項?には, 7つの選択肢が存在する.
 - 温度=暑い: 0/2
 - 温度=暖かい: 2/2
 - 温度=涼しい: 1/1
 - 湿度=高い: 1/3
 - 湿度=普通: 2/2
 - 風=有り: 1/2
 - 風=無し: 2/3
- 「温度=暖かい」と「温度=涼しい」が最大正答率100%で相互に独立なので, 以下のルールを得る.
 - If [天気=晴れ] and [温度=暖かい] Then ハイキング=○
 - If [天気=晴れ] and [温度=涼しい] Then ハイキング=○
- これらのルールは3事例のうちの3事例をカバーしているので, 終了.

21

問題3: 解答例

- 以上より, 以下のルールを得る.
 - If [天気=曇り] and [湿度=普通] Then ハイキング=○
 - If [天気=晴れ] and [温度=暖かい] Then ハイキング=○
 - If [天気=晴れ] and [温度=涼しい] Then ハイキング=○

22