

14.5 강의 평가점수 Statistics.com의 통계교육원(The Institute for Statistics Education)에서는 학생들이 한 과목을 수료하는 즉시 수료한 과목에 대해 다양한 측면에서 평가하도록 한다. 이 교육원에서는 평가가 끝난 후 학생들에게 추가적인 과목을 추천하는 시스템을 고려하고 있다. <표 14.16>에 나타난 학생들의 일부 온라인 강의 평가점수와 E.N. 학생에게 어떤 과목을 추가적으로 추천할지 고려하시오,

```
setwd("c:/rdata")
rating <- read.csv("c:/rdata/courserating.csv")
rating[,1]
row.names(rating) <- rating[,1]
row.names(rating)

m <- as.matrix(rating[, -1])
mm <- m[c(1,4,15),]
library(recommenderlab)
r <- as(m, "realRatingMatrix")
rr <- as(mm, "realRatingMatrix")
```

a. 먼저 사용자 기반 협업필터를 고려해보자. 이는 모든 학생들 간의 상관계수의 계산을 요구한다. 어떤 학생들에 대해 E.N. 학생과의 상관계수를 계산할 수 있는가? 이를 계산하시오.

$$\bar{r}_{LN} = (4 + 3 + 2 + 4 + 2) / 5 = 3$$

$$\bar{r}_{DS} = (4 + 2 + 4) / 3 = 3.33$$

$$\bar{r}_{EN} = (4 + 4 + 4 + 3) / 4 = 3.75$$

$$\text{Corr}(\bar{r}_{LN}, \bar{r}_{EN}) = \frac{(4-3)(4-3.75) + (4-3)(4-3.75) + (2-3)(3-3.75)}{\sqrt{(4-3)^2 + (4-3)^2 + (2-3)^2} \sqrt{(4-3.75)^2 + (4-3.75)^2 + (3-3.75)^2}} = 0.87$$

$$\text{Corr}(\bar{r}_{DS}, \bar{r}_{EN}) = \frac{(4-3.33)(4-3.75) + (2-3.33)(4-3.75) + (4-3.33)(4-3.75)}{\sqrt{(4-3.33)^2 + (2-3.33)^2 + (4-3.33)^2} \sqrt{(4-3.75)^2 + (4-3.75)^2 + (4-3.75)^2}} = 0$$

b. E.N. 학생과 가장 가까운 학생을 바탕으로 E.N. 학생에게 어떤 과목을 추천해야 하겠는가? 이유를 설명하시오.

```
UB.Rec.pea <- Recommender(r, "UBCF", parameter="pearson")
pred <- predict(UB.Rec.pea, r, type="ratings")
pmmm <- as(pred, "matrix")
pmmm
```

	SQL	Spatial	PA1	DM.in.R	Python	Forecast	R.Prog	Hadoop	Regression
LN	NA	NA	NA	2.703641	NA	NA	NA	NA	NA
MH	NA	NA	NA	3.388491	NA	NA	4.131714	NA	3.244415
JH	NA	NA	NA	2.000000	2.000000	NA	2.000000	NA	2.000000
EN	NA	NA	NA	NA	3.866128	NA	NA	NA	NA
DU	NA	NA	NA	4.000000	4.000000	NA	4.000000	NA	4.000000
FL	NA	NA	NA	NA	NA	NA	NA	NA	NA
GL	NA	NA	NA	NA	NA	NA	NA	NA	NA
AH	NA	NA	NA	NA	NA	NA	NA	NA	NA
SA	NA	NA	NA	NA	NA	NA	NA	4	NA
RW	NA	NA	NA	NA	NA	NA	NA	NA	NA
BA	NA	NA	NA	NA	NA	NA	NA	4	NA
MG	NA	NA	NA	NA	NA	NA	NA	4	NA
AF	NA	NA	NA	NA	NA	NA	NA	4	NA
KG	NA	NA	NA	NA	NA	NA	NA	3	NA
DS	NA	NA	NA	NA	3.444444	NA	NA	NA	3.000000

>>E.N. 학생에게 Python과목을 추천한다.

이유는 E.N.의 교과목 rating 예측값이 3.866128인 과목을 추천한다.(rating을 사용해서 이미 들을 과목은 추천하지 말아라)

c. R의 Similarity() 함수를 이용하여 E.N.과 고객 간 코사인 유사도를 계산하시오.

(co_rated가 하나라면 유사도 계산이 의미가 없음)

```
similarity(rr, method = "cosine")
```

```
      LN      EN  
EN 0.9891005  
DS 1.0000000 0.9622504
```

>> E.N.과 L.N.의 코사인 유사도는 0.9891005

>> E.N.과 D.S.의 코사인 유사도는 0.9622504

d. E.N 학생과 코사인 유도상으로 가장 가까운 학생을 바탕으로 E.N. 학생에게 어떤 학업과정을 추천해야 하겠는가?

```
UB.Rec.cos <- Recommender(r, "UBCF")  
pred <- predict(UB.Rec.cos, r, type="ratings")  
cmmm <- as(pred, "matrix")  
cmmm
```

	SQL	Spatial	PA1	DM.in.R	Python	Forecast	R.Prog	Hadoop	Regression
LN	NA	NA	NA	2.703641	NA	NA	NA	NA	NA
MH	NA	NA	NA	3.388491	NA	NA	4.131714	NA	3.244415
JH	NA	NA	NA	2.000000	2.000000	NA	2.000000	NA	2.000000
EN	NA	NA	NA	NA	3.866128	NA	NA	NA	NA
DU	NA	NA	NA	4.000000	4.000000	NA	4.000000	NA	4.000000
FL	NA	NA	NA	NA	NA	NA	NA	NA	NA
GL	NA	NA	NA	NA	NA	NA	NA	NA	NA
AH	NA	NA	NA	NA	NA	NA	NA	NA	NA
SA	NA	NA	NA	NA	NA	NA	NA	4	NA
RW	NA	NA	NA	NA	NA	NA	NA	NA	NA
BA	NA	NA	NA	NA	NA	NA	NA	4	NA
MG	NA	NA	NA	NA	NA	NA	NA	4	NA
AF	NA	NA	NA	NA	NA	NA	NA	4	NA
KG	NA	NA	NA	NA	NA	NA	NA	3	NA
DS	NA	NA	NA	NA	3.444444	NA	NA	NA	3.000000

>>E.N. 학생에게 Python 과목을 추천한다.

이유는 E.N.의 교과목 rating 예측값이 3.866128인 과목을 추천한다.

g. 항목 기반 협업 필터링이 이 데이터에 적용하고(R을 이용), 그 결과를 기반으로 E.N. 학생에게 과목을 추천하시오.

```
IB.Rec <- Recommender(r, "IBCF")
ipred <- predict(IB.Rec, r, n=5, type = "topNList")
immm <- as(ipred, "matrix")
immm
```

	SQL	Spatial	PA1	DM.in.R	Python	Forecast	R.Prog	Hadoop	Regression
LN	NA	4	2	3.26231	NA	NA	NA	NA	NA
MH	NA	NA	NA	3.00000	NA	3.000000	3	NA	3.000000
JH	NA	NA	NA	NA	NA	NA	NA	NA	NA
EN	NA	4	NA	NA	4	3.666667	NA	NA	NA
DU	NA	NA	NA	NA	NA	NA	NA	NA	NA
FL	NA	NA	NA	NA	NA	NA	NA	NA	NA
GL	NA	NA	NA	NA	NA	NA	NA	NA	NA
AH	NA	NA	NA	NA	NA	NA	NA	NA	NA
SA	NA	NA	NA	NA	NA	NA	NA	NA	NA
RW	NA	NA	NA	NA	NA	2.000000	NA	NA	NA
BA	NA	NA	NA	NA	NA	NA	NA	NA	NA
MG	NA	NA	NA	NA	NA	NA	NA	NA	NA
AF	NA	NA	NA	NA	NA	NA	NA	NA	NA
KG	NA	NA	NA	NA	NA	NA	NA	NA	NA
DS	NA	4	NA	NA	4	4.000000	NA	NA	3.296585

>> E.N. 학생에게 Spatial > Python > Forecast을 추천한다.