Learning Recommendation System

Create a personalized study path for students

Eddie Lin

2018/03/15

Project Description

Recommendation system is everywhere in our lives. Amazon, YouTube, Spotify, you name it. On the other hand, how can we use recommendation system in education? As a learning recommendation system can become very complicated with algorithms and parameters, this project will demonstrate a simpler system that suggest a student's learning path. By pooling data from students in the same class (like pooling preference from millions of YouTube viewers), this recommendation system can suggest a student which study unit to take and in their sequence. Some interesting to be explored are:

1. How does this learning recommendation system generate different paths for individual students?

2. Do those learning paths make sense from a pedagogical perspective? Especially regarding the arrangement of units and curriculum designs?

3. Will students really follow the suggested paths to study each unit?

Tools & Data

- R
- R packages: dplyr, tidyr, lsa
- Data source: a student survey from a master's program at a private university

Data Description

The data comes from a in-class survey asking students about their interest and perceived difficulty of 6 different study units: (1)Prediction, (2)Social Networks, (3)Neural Networks, (4)Natural Language Processing, (5)Interactive Visualization, (6)Turning Data Insights into Actions

Students were asked to answer the following 2 questions about each of those study units. The students have previewed the introductions of each study unit before they answered these 2 questions.

1. How much are you interested in this study unit?

2. How difficult do you think this study unit is to you?

Load & preprocess the data

- the survey asked students about the the 6 units in terms of both interest and difficult
- We will be using matrix operations in this assignment, so will have to convert data frames to matrices

```
library(dplyr)
library(tidyr)
I1 <- read.csv("interest.csv")
D1 <- read.csv("difficulty.csv")

I2 <- select(I1, 2:7)
I2 <- as.matrix(I2)
rownames(I2) <- I1$stid
D2 <- select(D1, 2:7)
D2 <- as.matrix(D2)
rownames(D2) <- D1$stid
```

Students' interest data

- We can generate a user-based similarity matrix based on cosine similarity using the ratings
- This matrix will represent the similarity of interests between students in the class.
- We will transpose the matrix so that multiplication occurs by students rather than units

```
I2 <- t(I2)

library(lsa)
I.SIM <- cosine(I2)
diag(I.SIM) <- NA # because students will be most similar to themsel
```

Find out which student is similar to you

- We can find the first few students whose interest are most similar to our designated student
- In practice, we could also combine a few questions and students' responses so we can develop a more complex interest matrix
- Doing this could be useful to break students into different work groups. Either based on similar or dissimilar interests

```
student.one <- "bgr2106" #Input your UNI id in here
head(rownames(I.SIM[order(I.SIM[student.one,], decreasing = TRUE),]),
```

```
## [1] "sz2686"  "rc3153"  "xw2509"  "sc4071"  "scd4123"
```

Suggest next study unit for students

- On the other hand, we can also use students' responses to difficulty and create another similarity matrix to suggest next unit to study
- This suggestion is based on the whole student body and their rankings of all 6 units regarding their perceived difficulty
- If someone decides to begin with the "prediction" unit, what will be the path for that person?

```
D.SIM <- cosine(D2)
diag(D.SIM) <- NA
head(rownames(D.SIM[order(D.SIM["pred.dif",], decreasing = TRUE),]),
```

```
## [1] "neural.dif" "nlp.dif"    "loop.dif"   "sna.dif"    "viz.dif"
```

Make sure students don't just study the easy stuff

- In teaching & learning, sometimes we can't let students choose their own study path solely based on their interest because they may only pick stuff that's easy
- We can avoid (as much) this by creating a composite measure that factor in different features at the same time to balance out students' choice biases
- In this current case, we will use Principal Component Analysis(PCA) to create another similartiy matrix and give suggestion about next unit after the "prediction unit"

Similarity matrix based on PCA

```
I3 <- gather(I1,unit,interest, 2:7)
D3 <- gather(D1, stid, difficulty)
C1 <- data.frame(I3$stid, I3$unit, I3$interest, D3$difficulty)
names(C1) <- c("stid", "unit", "interest", "difficulty")
C1 <- filter(C1, difficulty > 0)
C2 <- select(C1, "interest", "difficulty")
pc <- prcomp(C2)
C3 <- data.frame(C1$stid, C1$unit, pc$x)
C4 <- select(C3, C1.stid, C1.unit, PC1)
#Remove int from unit label
C4$C1.unit <- gsub(".int", "", C4$C1.unit)
C5 <- spread(C4, C1.stid, PC1)
row.names(C5) <- C5$C1.unit
C5$C1.unit <- NULL
C5 <- as.matrix(C5)
C5 <- ifelse(is.na(C5), 0, C5)
C5 <- t(C5)
C.SIM <- cosine(C5)
diag(C.SIM) <- NA
head(rownames(C.SIM[order(C.SIM["pred",], decreasing = TRUE),]), n =
```

Wait, what are our research questions again ???

1.**How does this learning recommendation system generate different paths for individual students?**

- We can use students' interest, perceoved difficulty, or a combination of different features to create a composite measure and give a suggeston for study path using cosine similarity

2.**Do those learning paths make sense in from a pedagogicall perspective? Especially regarding the arrangement of units and curriculum designs?**

- Yes and no. The recommendation system suggests a path that student could go by. This may fit most students' interest and perceived difficulty but they may not know how skills from one unit can/should add onto another.

3.**Will students really follow the suggested paths to study each unit?**

- Will customers buy everything that is recommended by Amazon? Not really. Sometimes students may begin with one unit and then realize things are not like what they thought. They may feel some units are particularly hard and will take some easier unit later on before another challenging unit

Findings & Summary

1.We found that different suggested paths based on students' interest and on a composite measure (interest + difficulty)

2.By using a composite measure, we can better prevent the problem of "letting students do the way the want but not learning much"

3.The suggested paths may not make perfect sense to the instructors, as their teaching experience could give other suggestion about how to scaffold knowledge through different units and what is the similar knowledge underlying some units.

Limitations & Suggestions

No data analytics is perfect, I came up with a few thoughts and make some suggestions in the followings:

1. In the current case, the surveyed students may not have a good understanding of each unit by just previewing the unit introduction, so their resonses to the **interest** and the **difficulty** questions may be biased.

2. The survey only has 2 questions on 5-point Likert scale and was given to 22 students , the validity of the results may be increased by increasing the sameple size and variety of questions. But most importantly, we have to make sure students do understand those units before they give responses.

Limitations & Suggestions (continued)

3. A good learning recommendation system, just like any recommendation system, takes time and data to train and revise. In practice, we could probably use this technique on some online courses where data and users' feedback are more easily collected. Ideally, we will use data from the past courses and many features to create a composite measure to compute users' similarity. From there, we can give new student some recommendation for their study path based on similarity of choices.

4. Arguably, longer model training time & more features from users can help a learning recommendation system to improve. But in education, we have to be very careful if the model makes pedagogical sense and make sure students are constantly making progress, so they don't indulge themselves in their local maximum where they just know things they already knew before