# hw1_643_DieudonneO

*Dieudonne*

*June 19, 2016*

## RECOMMENDER SYSTEM ON MOVIE LENS DATA

## INTRODUCTION

***This is a mini project I wrote for my course Data 643 at CUNY***

*The goal here is to explore recommenderlab,write few functions and predict recommendations to users based on the User-Based -Collaborative-Filtering(UBCF) which through this project appears to be the most suited*

There are 2 sets of data u.data which is ratings data and u.item data which is movie data

The data are located here http://grouplens.org/datasets/movielens/

```
#RECOMMENDER SYSTEM ON MOVIE LENS DATA

#(The data are located here http://grouplens.org/datasets/movielens/)
library(recommenderlab)
library(reshape2)
```

## FUNCTION TO GRAB THE DATA

```
get.Data <- function(){

  ##laod ratings data
  ratings <- read.delim("~/Downloads/u.data.txt", header=F)
  colnames(ratings) <- c("userID","movieID","rating", "timestamp")

  ## load movies data
  movies <- read.delim("~/Downloads/u.item.txt", sep="|", header=F, stringsAsFactors = FALSE)
  colnames(movies)[colnames(movies)=="V1"] <- "movieID"
  colnames(movies)[colnames(movies)=="V2"] <- "name"

  return(list(ratings=ratings, movies=movies))

}
```

## FUNCTION FOR DATA PREPARATION AND PROCESSING

```
Pre.Process = function(ratings, movies)
{
  ratings[,2] <- dataList$movies$name[as.numeric(ratings[,2])]
```

```
  # remove duplicate entries for any user-movie combination
  ratings <- ratings[!duplicated(ratings[,1:2]),]
}
```

## Function to Create movie ratingMatrix from rating Data and movie data

```
Create.Rating.Matrix <- function(ratings)
{
  # converting the ratingData data frame into rating matrix
  Ratings.Mat <- dcast( ratings, userID ~ movieID, value.var = "rating" , index="userID")
  ratings <- Ratings.Mat[,2:ncol(Ratings.Mat)]

  Ratings.Mat.Fin <- as(ratings, "matrix")  ## cast data frame as matrix
  movie.Rating.Mat <- as(Ratings.Mat.Fin, "realRatingMatrix")   ## create the realRatingMatrix
  ### setting up the dimnames ###
  dimnames(movie.Rating.Mat)[[1]] <- row.names(ratings)
  return (movie.Rating.Mat)
}
```

# MODELS

```
evaluateModels <- function(movie.Rating.Mat)

{
  ## Find out and analyse available  recommendation algorithm options for realRatingMatrix data
  recommenderRegistry$get_entries(dataType = "realRatingMatrix")

  scheme <- evaluationScheme(movie.Rating.Mat, method = "split", train = .9,
                             k = 1, given = 10, goodRating = 4)

  algorithms <- list(
    RANDOM = list(name="RANDOM", param=NULL),
    POPULAR = list(name="POPULAR", param=NULL),
    UBCF = list(name="UBCF", param=NULL)
  )

  # run algorithms, predict next n movies
  res <- evaluate(scheme, algorithms, n=c(1, 3, 5, 10, 15, 20))

  ## select the first results

  return (res)
}
```

# VISUALIZATION

```
graphs <- function(res)
{
  # Draw ROC curve
  plot(res, annotate = 1:3, legend="topright")

  # See precision / recall
  plot(res, "prec/rec", annotate=3, legend="topright", xlim=c(0,.22))
}
```

# CREATE FUNCTION FOR PREDICTION MODEL

```
create.Model <-function (movie.Rating.Mat,method){

  model <- Recommender(movie.Rating.Mat, method = method)
  names(getModel(model))
  getModel(model)$method

  getModel(model)$nn

  return (model)
}
```

### RATINGS PREDICTIONS USING USER BASED C FILTERING RECOMMENDATIONS

```
rec <- function(movie.Rating.Mat, model, userID, n)
{

  ### PREDICT THE TOP N recommendations for given user
  Top.N.List <-predict(model,movie.Rating.Mat[userID],n=n)
  as(Top.N.List,"list")
}
```

# LOAD MOVIE LENS DATA

```
dataList<- get.Data()
```

# DATA PREPARATION AND PROCESSING

```
ratings<- Pre.Process(dataList$ratings, dataList$movies)
```

## MATRIX FOR MOVIE RATING

```
movie.Rating.Mat<- Create.Rating.Matrix(ratings)
```

## MODELS EVALUTION

```
evalList <- evaluateModels(movie.Rating.Mat)
```

```
## RANDOM run fold/sample [model time/prediction time]
##   1  [0.004sec/0.418sec]
## POPULAR run fold/sample [model time/prediction time]
##   1  [0.015sec/0.085sec]
## UBCF run fold/sample [model time/prediction time]
##   1  [0.01sec/1.299sec]
```

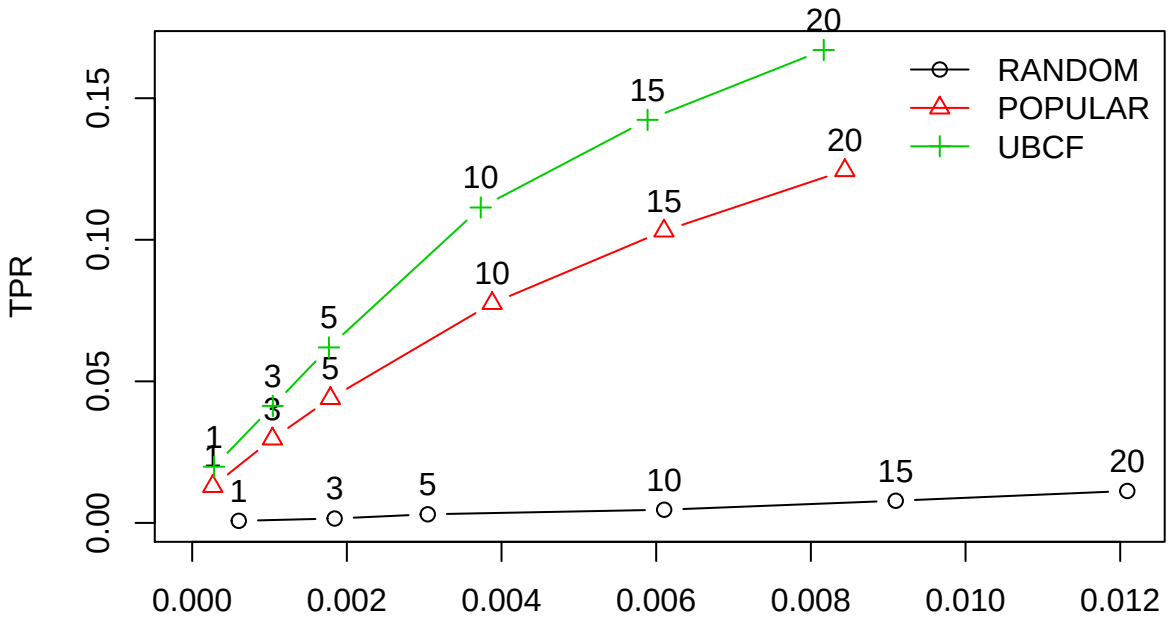
```
evalList
```

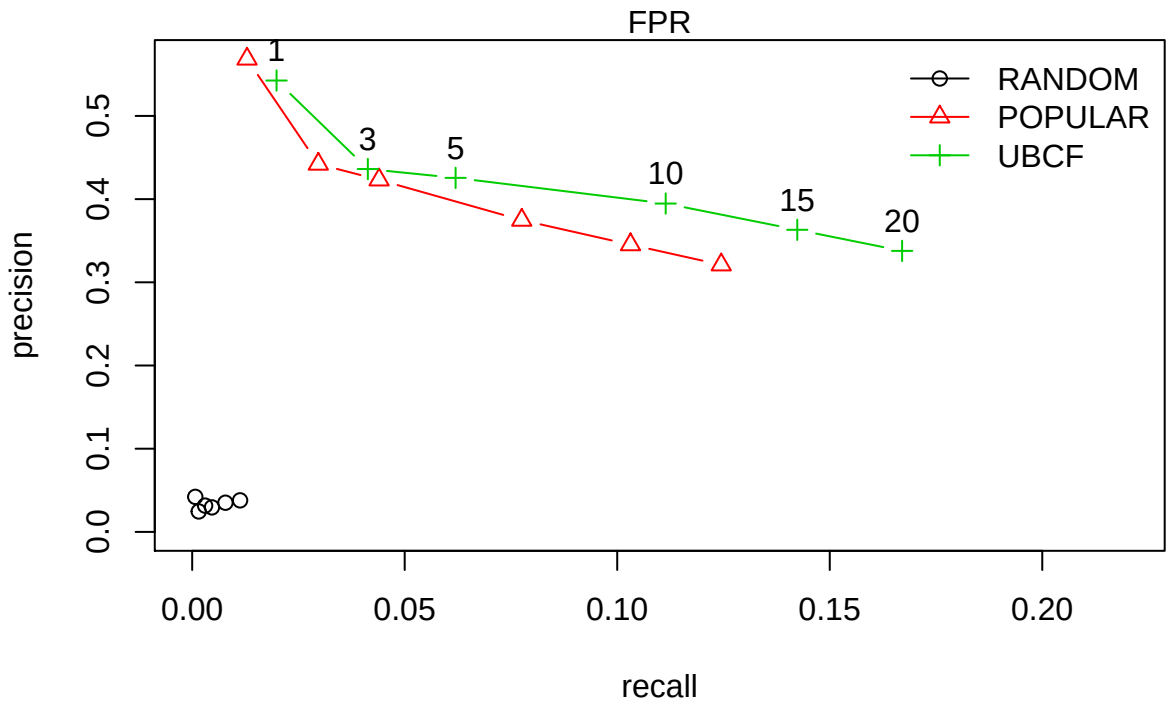
```
## List of evaluation results for 3 recommenders:
## Evaluation results for 1 folds/samples using method 'RANDOM'.
## Evaluation results for 1 folds/samples using method 'POPULAR'.
## Evaluation results for 1 folds/samples using method 'UBCF'.
```

### The plot for comparing "Random", "Popular", "UBCF" based recommender algorithm is shown:

### plot evaluation result

```
graphs(evalList)
```


TPR
FPR
RANDOM
POPULAR
UBCF
0.00
0.05
0.10
0.15
0.000
0.002
0.004
0.006
0.008
0.010
0.012
1
3
5
10
15
20


precision
recall
RANDOM
POPULAR
UBCF
0.0
0.1
0.2
0.3
0.4
0.5
0.00
0.05
0.10
0.15
0.20
1
3
5
10
15
20

**on visualization, looks like UBCF has highest precision.**

## The visualisation shows "UBCF" algorithm has highest precision. So I picked "UBCF" to predicts top 10 recommendation of user with userID = 1.

## get Confusion matrix for "UBCF"

```
getConfusionMatrix(evalList[["UBCF"]])[[1]][,1:4]
```

```
##           TP         FP       FN       TN
## 1  0.5368421  0.4526316 62.63158 1590.379
## 3  1.2947368  1.6736842 61.87368 1589.158
## 5  2.1052632  2.8421053 61.06316 1587.989
## 10 3.9052632  5.9894737 59.26316 1584.842
## 15 5.3894737  9.4526316 57.77895 1581.379
## 20 6.6842105 13.1052632 56.48421 1577.726
```

### LET DO THE RECOMMENDATION BASED ON "UBCF"

```
rec_model <- create.Model(movie.Rating.Mat, "UBCF")
userID <- 1
topN <- 5
rec(movie.Rating.Mat, rec_model, userID, topN)
```

```
## [[1]]
## [1] "Glory (1989)"             "Schindler's List (1993)"
## [3] "Close Shave, A (1995)"    "Casablanca (1942)"
## [5] "Leaving Las Vegas (1995)"
```

```
userID<-2
topN<-10
rec(movie.Rating.Mat, rec_model, userID, topN)
```

```
## [[1]]
##  [1] "Lone Star (1996)"                "Boot, Das (1981)"
##  [3] "Dead Man Walking (1995)"         "Celluloid Closet, The (1995)"
##  [5] "Return of the Jedi (1983)"       "Casablanca (1942)"
##  [7] "Angels and Insects (1995)"       "Breaking the Waves (1996)"
##  [9] "Seven Years in Tibet (1997)"     "Welcome to the Dollhouse (1995)"
```

## Let recommend the top 10 movies for users with ID between 5 and 15

```
for (userID in 5:15){
  print("We recommend you those movies")
  print(rec(movie.Rating.Mat,rec_model,userID,topN))
}
```

```
## [1] "We recommend you those movies"
## [[1]]
##  [1] "Terminator 2: Judgment Day (1991)"
##  [2] "Terminator, The (1984)"
##  [3] "Usual Suspects, The (1995)"
##  [4] "Contact (1997)"
##  [5] "Braveheart (1995)"
##  [6] "Casablanca (1942)"
##  [7] "Twelve Monkeys (1995)"
##  [8] "Godfather, The (1972)"
##  [9] "Shawshank Redemption, The (1994)"
## [10] "Raising Arizona (1987)"
##
## [1] "We recommend you those movies"
## [[1]]
##  [1] "Empire Strikes Back, The (1980)" "Rear Window (1954)"
##  [3] "Chinatown (1974)"                "Clockwork Orange, A (1971)"
##  [5] "Singin' in the Rain (1952)"      "Return of the Jedi (1983)"
##  [7] "Ran (1985)"                      "Titanic (1997)"
##  [9] "All About Eve (1950)"            "High Noon (1952)"
##
## [1] "We recommend you those movies"
## [[1]]
##  [1] "Lone Star (1996)"           "Miller's Crossing (1990)"
##  [3] "Hoop Dreams (1994)"         "Leaving Las Vegas (1995)"
##  [5] "Big Night (1996)"           "Close Shave, A (1995)"
##  [7] "Titanic (1997)"             "This Is Spinal Tap (1984)"
##  [9] "Wrong Trousers, The (1993)" "Quiz Show (1994)"
##
## [1] "We recommend you those movies"
## [[1]]
##  [1] "Titanic (1997)"
##  [2] "Shawshank Redemption, The (1994)"
##  [3] "Usual Suspects, The (1995)"
##  [4] "Silence of the Lambs, The (1991)"
##  [5] "Fargo (1996)"
##  [6] "L.A. Confidential (1997)"
##  [7] "Schindler's List (1993)"
##  [8] "Bridge on the River Kwai, The (1957)"
##  [9] "Boot, Das (1981)"
## [10] "Good Will Hunting (1997)"
##
## [1] "We recommend you those movies"
## [[1]]
##  [1] "Air Force One (1997)"           "Contact (1997)"
##  [3] "Titanic (1997)"                 "Raiders of the Lost Ark (1981)"
##  [5] "Wag the Dog (1997)"             "Scream (1996)"
##  [7] "Good Will Hunting (1997)"       "Apt Pupil (1998)"
```

```
##  [9] "L.A. Confidential (1997)"       "Apostle, The (1997)"
##
## [1] "We recommend you those movies"
## [[1]]
##  [1] "Killing Fields, The (1984)"
##  [2] "Godfather: Part II, The (1974)"
##  [3] "High Noon (1952)"
##  [4] "Empire Strikes Back, The (1980)"
##  [5] "Schindler's List (1993)"
##  [6] "Blade Runner (1982)"
##  [7] "To Kill a Mockingbird (1962)"
##  [8] "Mr. Smith Goes to Washington (1939)"
##  [9] "Great Escape, The (1963)"
## [10] "My Fair Lady (1964)"
##
## [1] "We recommend you those movies"
## [[1]]
##  [1] "Titanic (1997)"                   "Good Will Hunting (1997)"
##  [3] "L.A. Confidential (1997)"         "Star Wars (1977)"
##  [5] "Godfather, The (1972)"            "Shawshank Redemption, The (1994)"
##  [7] "Trainspotting (1996)"             "Raiders of the Lost Ark (1981)"
##  [9] "As Good As It Gets (1997)"        "Return of the Jedi (1983)"
##
## [1] "We recommend you those movies"
## [[1]]
##  [1] "To Kill a Mockingbird (1962)"
##  [2] "Shawshank Redemption, The (1994)"
##  [3] "Braveheart (1995)"
##  [4] "Casablanca (1942)"
##  [5] "Toy Story (1995)"
##  [6] "Indiana Jones and the Last Crusade (1989)"
##  [7] "One Flew Over the Cuckoo's Nest (1975)"
##  [8] "Great Escape, The (1963)"
##  [9] "Fargo (1996)"
## [10] "Sling Blade (1996)"
##
## [1] "We recommend you those movies"
## [[1]]
##  [1] "Citizen Kane (1941)"
##  [2] "It's a Wonderful Life (1946)"
##  [3] "Unforgiven (1992)"
##  [4] "Raging Bull (1980)"
##  [5] "Vertigo (1958)"
##  [6] "Mr. Smith Goes to Washington (1939)"
##  [7] "Fried Green Tomatoes (1991)"
##  [8] "Third Man, The (1949)"
##  [9] "Gone with the Wind (1939)"
## [10] "Killing Fields, The (1984)"
##
## [1] "We recommend you those movies"
## [[1]]
##  [1] "Casablanca (1942)"
##  [2] "Citizen Kane (1941)"
##  [3] "Chasing Amy (1997)"
```

```
##  [4] "My Life as a Dog (Mitt liv som hund) (1985)"
##  [5] "Wizard of Oz, The (1939)"
##  [6] "Third Man, The (1949)"
##  [7] "Richard III (1995)"
##  [8] "Eat Drink Man Woman (1994)"
##  [9] "Vertigo (1958)"
## [10] "Babe (1995)"
##
## [1] "We recommend you those movies"
## [[1]]
##  [1] "Fargo (1996)"
##  [2] "Willy Wonka and the Chocolate Factory (1971)"
##  [3] "Titanic (1997)"
##  [4] "Boot, Das (1981)"
##  [5] "Amistad (1997)"
##  [6] "Good Will Hunting (1997)"
##  [7] "Leaving Las Vegas (1995)"
##  [8] "Close Shave, A (1995)"
##  [9] "Lone Star (1996)"
## [10] "Donnie Brasco (1997)"
```