hw1_643_DieudonneO

Dieudonne

June 19, 2016

RECOMMENDER SYSTEM ON MOVIE LENS DATA

INTRODUCTION

This is the second mini project I wrote for my course Data 643 at CUNY

I use mainly recommenderlab,write few functions and predict recommendations to users using various filtering methods and i compare the methods

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/

```
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),
    IBCF= list(name="IBCF",param=NULL),
    PCA=list(name="PCA",param=NULL),
    SVD=list(name="SVD",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 = 1:5, legend="topright")

  # See precision / recall
  plot(res, "prec/rec", annotate=5, 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

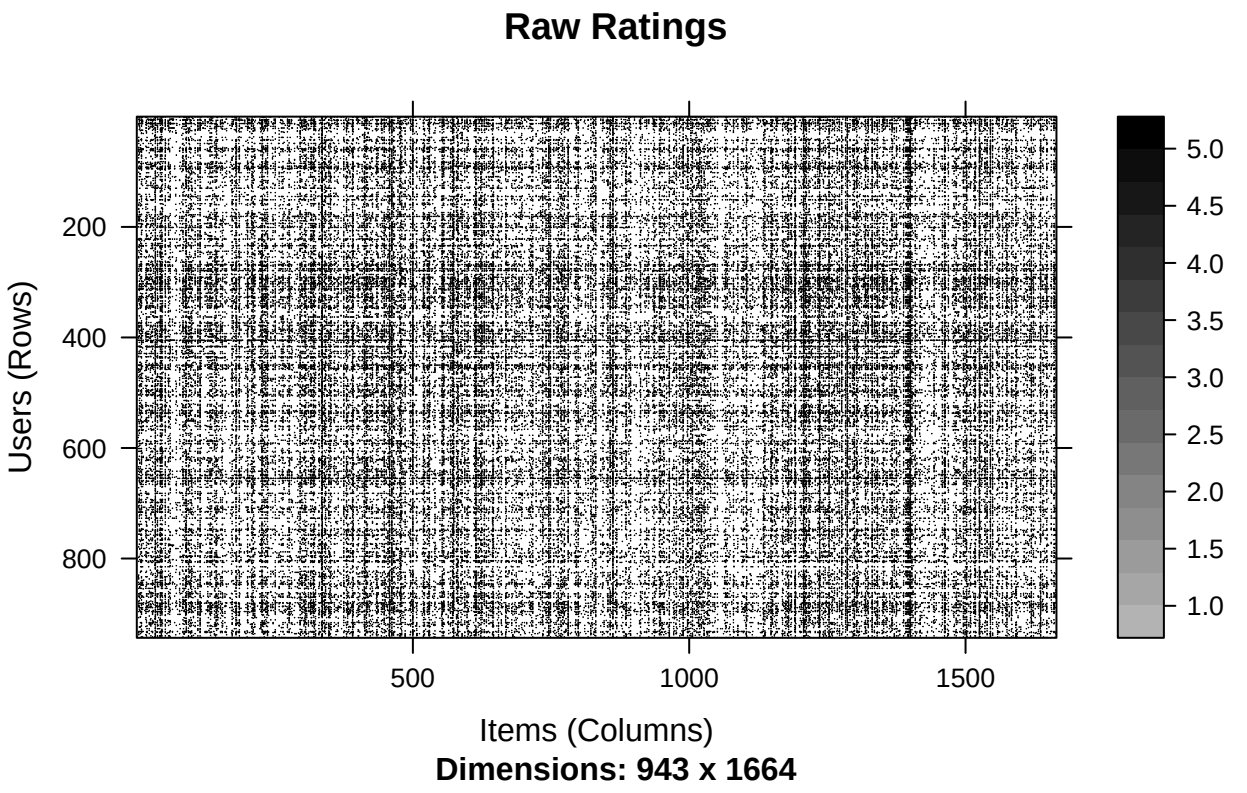
```
dataList<- get.Data()
```

DATA PREPARATION AND PROCESSING

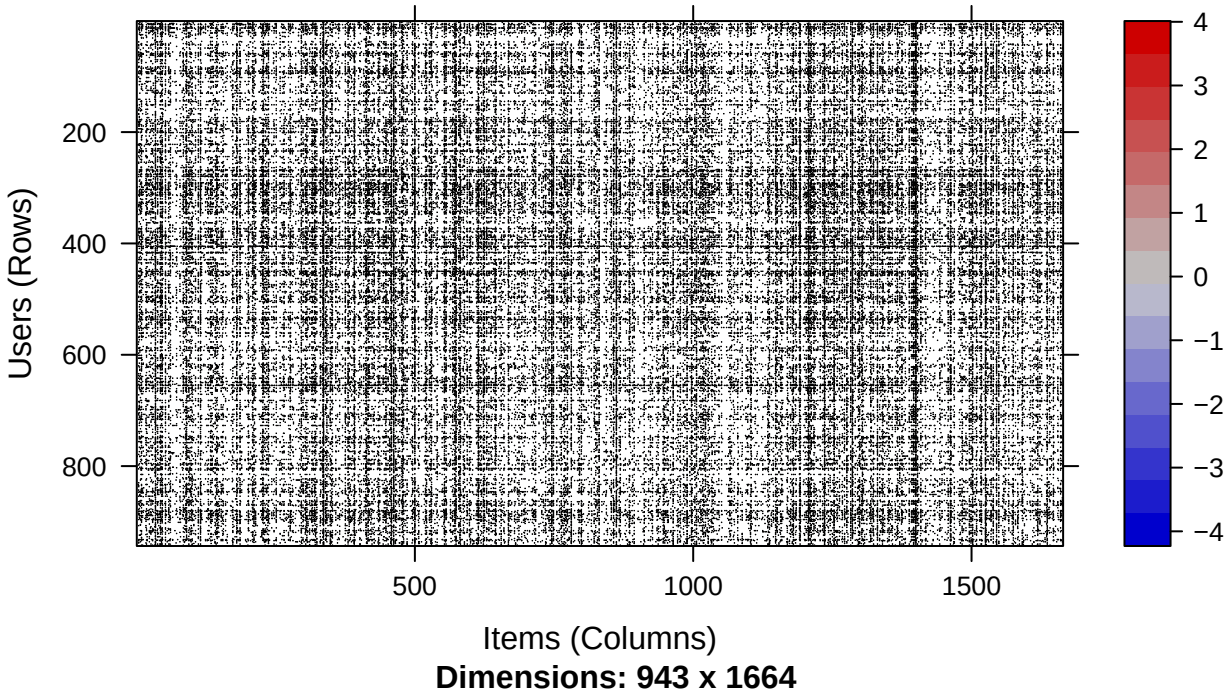
```
ratings<- Pre.Process(dataList$ratings, dataList$movies)
```

NORMALIZATION,BINARIZATION, REAL RATING MATRIX

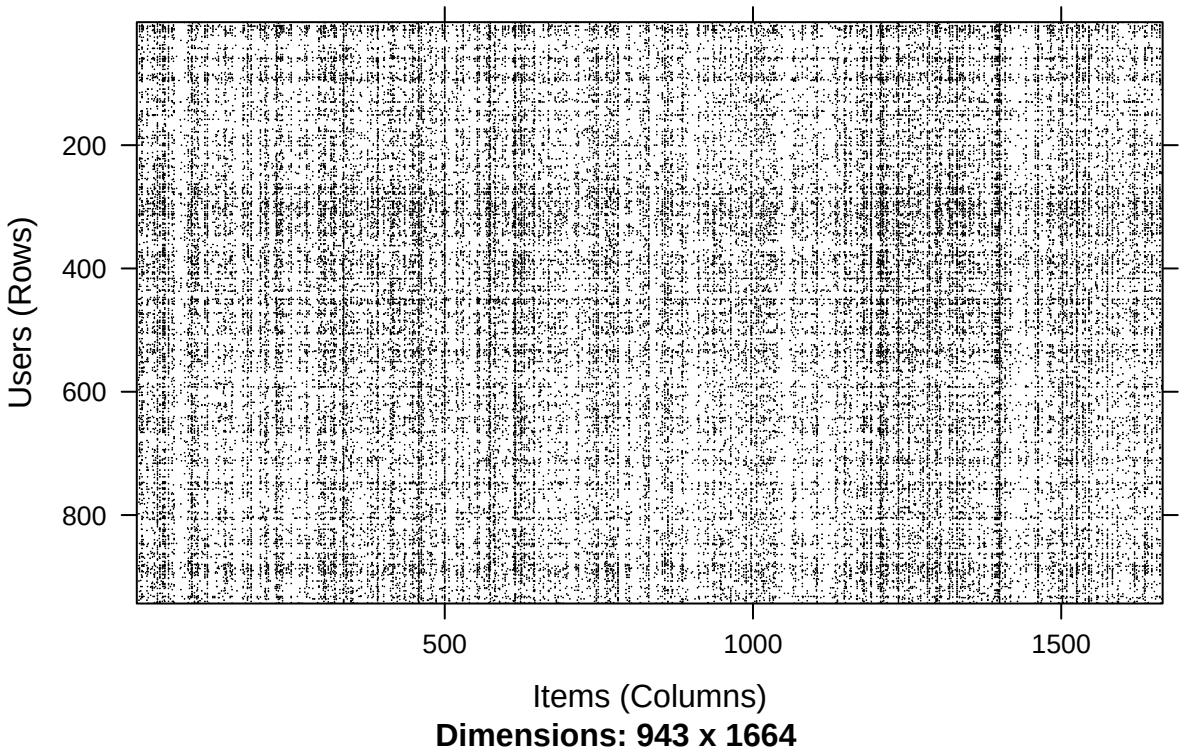
```
library(ggplot2)
library(Hmisc)
movie.Rating.Mat<- Create.Rating.Matrix(ratings)
l=as(movie.Rating.Mat,"list")
#str(l)

m<-as(movie.Rating.Mat,"matrix")
#head(m)
rm<-normalize(movie.Rating.Mat)
#str(rm)
#rm
#as(rm,"list")
image(movie.Rating.Mat,main="Raw Ratings")
```




```
image(rm,main="Normalized Ratings")
```




```
bm<-binarize(movie.Rating.Mat,minRating=4)
#str(bm)
image(bm,main="binarize data")
```



MODELS EVALUATION

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

```
## RANDOM run fold/sample [model time/prediction time]
##   1  [0.006sec/0.478sec]
## POPULAR run fold/sample [model time/prediction time]
##   1  [0.02sec/0.109sec]
## UBCF run fold/sample [model time/prediction time]
##   1  [0.009sec/1.652sec]
## IBCF run fold/sample [model time/prediction time]
##   1  [60.479sec/0.424sec]
## PCA run fold/sample [model time/prediction time]
##   1  Timing stopped at: 0.048 0.016 0.064
## SVD run fold/sample [model time/prediction time]
##   1  [0.009sec/14.257sec]

## Warning in .local(x, method, ...):
##   Recommender 'PCA' has failed and has been removed from the results!
```

```
evalList
```

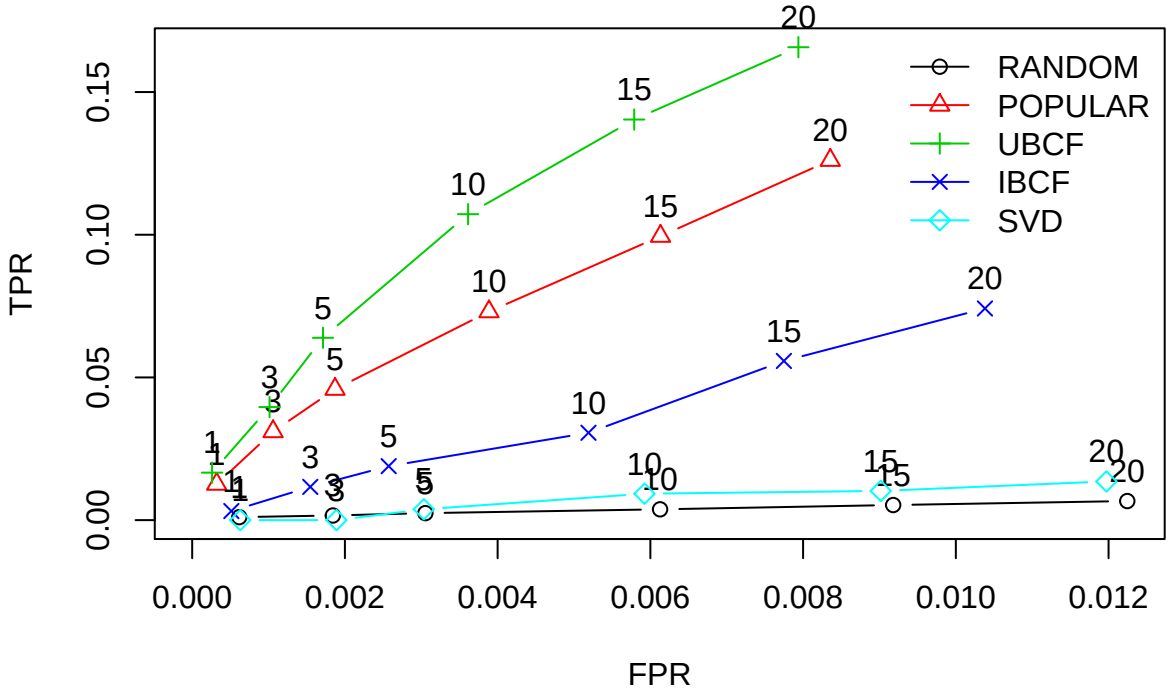
```
## List of evaluation results for 5 recommenders:
```

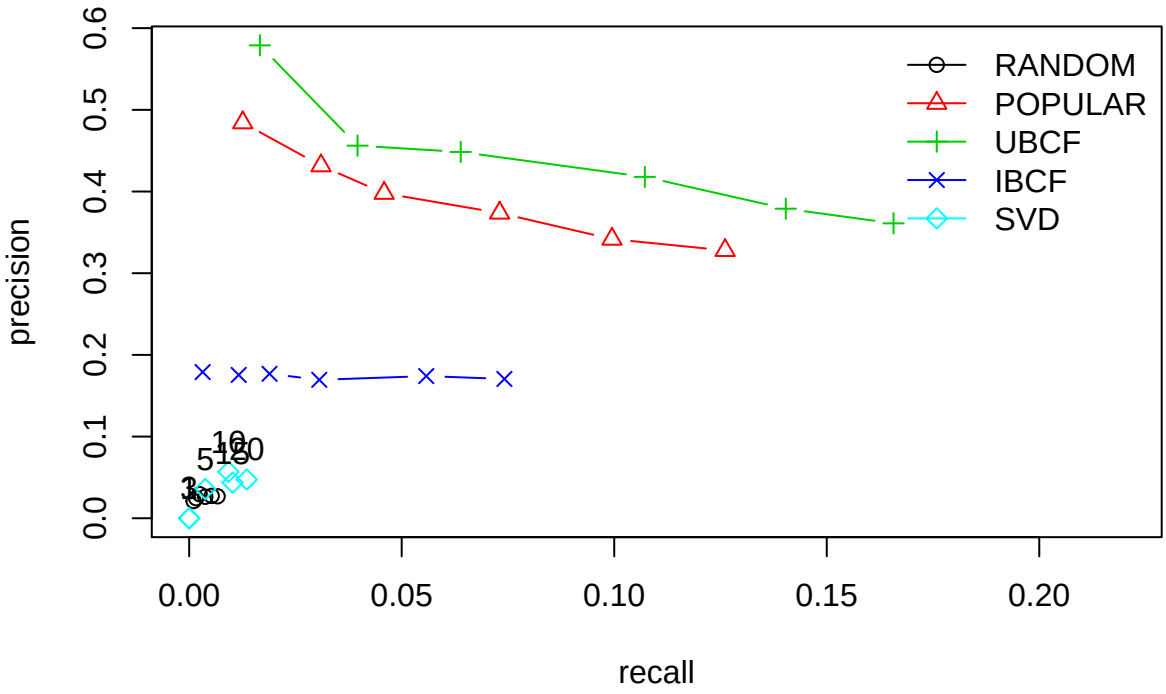
```
## Evaluation results for 1 folds/samples using method 'RANDOM'.
## Evaluation results for 1 folds/samples using method 'POPULAR'.
## Evaluation results for 1 folds/samples using method 'UBCF'.
## Evaluation results for 1 folds/samples using method 'IBCF'.
## Evaluation results for 1 folds/samples using method 'SVD'.
```

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

plot evaluation result

```
graphs(evalList)
```





CONCLUSION

CLEARLY UBCF got the better metrics compare to the other methods

CONFUSION MATRIX FOR ALL METHODS

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

```
##           TP         FP       FN       TN
## 1  0.5789474  0.4210526 63.04211 1589.958
## 3  1.3684211  1.6315789 62.25263 1588.747
## 5  2.2421053  2.7578947 61.37895 1587.621
## 10 4.1789474  5.8210526 59.44211 1584.558
## 15 5.6842105  9.3157895 57.93684 1581.063
## 20 7.2210526 12.7789474 56.40000 1577.600
```

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

```
##           TP         FP       FN       TN
## 1  0.1789474  0.8210526 63.44211 1589.558
## 3  0.5263158  2.4736842 63.09474 1587.905
## 5  0.8842105  4.1157895 62.73684 1586.263
## 10 1.6947368  8.3052632 61.92632 1582.074
## 15 2.6105263 12.3894737 61.01053 1577.989
## 20 3.4105263 16.5894737 60.21053 1573.789
```

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

```
##           TP         FP       FN       TN
## 1  0.4842105  0.5157895 63.13684 1589.863
## 3  1.2947368  1.7052632 62.32632 1588.674
## 5  1.9894737  3.0105263 61.63158 1587.368
## 10 3.7368421  6.2631579 59.88421 1584.116
## 15 5.1263158  9.8736842 58.49474 1580.505
## 20 6.5578947 13.4421053 57.06316 1576.937
```

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

```
##            TP         FP       FN       TN
## 1  0.02105263  0.9789474 63.60000 1589.400
## 3  0.07368421  2.9263158 63.54737 1587.453
## 5  0.14736842  4.8526316 63.47368 1585.526
## 10 0.26315789  9.7368421 63.35789 1580.642
## 15 0.41052632 14.5894737 63.21053 1575.789
## 20 0.53684211 19.4631579 63.08421 1570.916
```

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

```r
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)"
```

```
rec_model2 <- create.Model(movie.Rating.Mat, "IBCF")
userID <- 1
topN <- 5
rec(movie.Rating.Mat, rec_model2, userID, topN)
```

```
## [[1]]
## [1] "2 Days in the Valley (1996)"  "American in Paris, An (1951)"
## [3] "Basquiat (1996)"              "Boys, Les (1997)"
## [5] "Brassed Off (1996)"
```

```
userID<-2
topN<-10
rec(movie.Rating.Mat, rec_model2, userID, topN)
```

```
## [[1]]
##  [1] "12 Angry Men (1957)"          "2001: A Space Odyssey (1968)"
##  [3] "African Queen, The (1951)"    "Alien (1979)"
##  [5] "Aliens (1986)"                "Amadeus (1984)"
##  [7] "Apocalypse Now (1979)"        "Babe (1995)"
##  [9] "Back to the Future (1985)"    "Beautiful Thing (1996)"
```

```
rec_model3 <- create.Model(movie.Rating.Mat, "POPULAR")
userID <- 1
topN <- 5
rec(movie.Rating.Mat, rec_model3, userID, topN)
```

```
## [[1]]
## [1] "Schindler's List (1993)"
## [2] "Titanic (1997)"
## [3] "L.A. Confidential (1997)"
## [4] "Casablanca (1942)"
## [5] "One Flew Over the Cuckoo's Nest (1975)"
```

```
userID<-2
topN<-10
rec(movie.Rating.Mat, rec_model3, userID, topN)
```

```
## [[1]]
##  [1] "Raiders of the Lost Ark (1981)"   "Silence of the Lambs, The (1991)"
##  [3] "Schindler's List (1993)"          "Shawshank Redemption, The (1994)"
##  [5] "Empire Strikes Back, The (1980)"  "Return of the Jedi (1983)"
##  [7] "Usual Suspects, The (1995)"       "Casablanca (1942)"
##  [9] "Pulp Fiction (1994)"              "Princess Bride, The (1987)"
```

```
rec_model4 <- create.Model(movie.Rating.Mat, "RANDOM")
userID <- 1
topN <- 5
rec(movie.Rating.Mat, rec_model4, userID, topN)
```

```
## [[1]]
## [1] "Beyond Bedlam (1993)"
## [2] "Vegas Vacation (1997)"
## [3] "Stripes (1981)"
## [4] "Mouse Hunt (1997)"
## [5] "Butch Cassidy and the Sundance Kid (1969)"
```

```
userID<-2
topN<-10
rec(movie.Rating.Mat, rec_model4, userID, topN)
```

```
## [[1]]
##  [1] "Burnt By the Sun (1994)"  "Boogie Nights (1997)"
##  [3] "Some Mother's Son (1996)" "Perez Family, The (1995)"
##  [5] "Blue in the Face (1995)"  "Two Much (1996)"
##  [7] "Mighty, The (1998)"       "Fall (1997)"
##  [9] "Rocket Man (1997)"        "Star Maps (1997)"
```