```r
require(plyr)
require(reshape2)
setwd("/Users/sarpotd/Desktop/Coursera/Recommender Systems/week3/Assignment2/")
tags <- read.csv("data/movie-tags.csv", header=FALSE)
colnames(tags) <- c("movie_id", "tags")
ratings <- read.csv("data/ratings.csv", header=FALSE)
colnames(ratings) <- c("user_id", "movie_id", "ratings")
users <- read.csv("data/users.csv", header=FALSE)
colnames(users) <- c("user_id", "user_name")
movies <- read.csv("data/movie-titles.csv", header=FALSE)
colnames(movies) <- c("movie_id", "movie_name")


tags_unique <- unique(tags)
idf <- as.matrix(log(length(movies$movie_id)/table(tags_unique[,2])))

tfidf <- as.matrix(table(tags$tags,tags$movie_id))
tfidf_fun <- function(c) (tfidf[,c]*idf)
tfidf_colnames <- colnames(tfidf)
tfidf_rownames <- rownames(tfidf)
tfidf <- mapply(tfidf_fun,1:ncol(tfidf))
colnames(tfidf) <- tfidf_colnames
rownames(tfidf) <- tfidf_rownames
sqrt_sum <- function(c) (c/(sqrt(sum(c^2))))
norm_tfidf <- apply(tfidf,2,sqrt_sum)


temp_1 <- melt(ratings, id=c("user_id","movie_id"), measure="ratings")
temp_2 <- acast(temp_1, user_id ~ movie_id ~ variable)
user_ratings <- as.matrix(temp_2[,,1])


### Unweighted

#user_ratings <- ratings[ratings$ratings >= 3.5,]
#user_ratings <- as.matrix(table(user_ratings$movie_id,user_ratings$user_id))

user_ratings[is.na(user_ratings)] = 0
user_ratings_unweighted <- user_ratings
user_ratings_unweighted[user_ratings[,] >= 3.5] = 1
user_ratings_unweighted[user_ratings[,] < 3.5] = 0


user_vector <- norm_tfidf%*%t(user_ratings_unweighted)


d <- "4045"
user_vector_norm <- sqrt(sum(user_vector[,d]^2))
cosine_fun <- function(c) ((c*user_vector[,d]))
tfidf_colnames <- colnames(tfidf)
tfidf_rownames <- rownames(tfidf)
```

```r
test_user <- apply(norm_tfidf,2,cosine_fun)
colnames(test_user) <- tfidf_colnames
rownames(test_user) <- tfidf_rownames
test_user_sum <- apply(test_user,2,sum)
div_sum <- function(c) (test_user_sum[c]/(sqrt(sum(norm_tfidf[,c]^2))*user_vector_norm))
h <- mapply(div_sum,1:length(test_user_sum))
as.matrix(sort(h[!user_ratings[d,]],decreasing=TRUE))



### WEIGHTED

user_ratings_weighted <- user_ratings
mean_calc <- function(c) (mean(na.omit(c)))
user_ratings_weighted_mean <- as.matrix(apply(user_ratings_weighted,1,mean_calc))
user_ratings_weighted[is.na(user_ratings)] = 0

sub_mean <- function(c) (c - user_ratings_weighted_mean)
user_ratings_weighted_rownames <- rownames(user_ratings_weighted)
temp_1 <- apply(user_ratings_weighted,2,sub_mean)
rownames(temp_1) <- user_ratings_weighted_rownames
user_ratings_weighted <- temp_1
user_ratings_weighted[is.na(user_ratings)] = 0


user_vector <- norm_tfidf%*%t(user_ratings_weighted)


d <- "680"
user_vector_norm <- sqrt(sum(user_vector[,d]^2))
cosine_fun <- function(c) ((c*user_vector[,d]))
tfidf_colnames <- colnames(tfidf)
tfidf_rownames <- rownames(tfidf)
test_user <- apply(norm_tfidf,2,cosine_fun)
colnames(test_user) <- tfidf_colnames
rownames(test_user) <- tfidf_rownames
test_user_sum <- apply(test_user,2,sum)
div_sum <- function(c) (test_user_sum[c]/(sqrt(sum(norm_tfidf[,c]^2))*user_vector_norm))
h <- mapply(div_sum,1:length(test_user_sum))
as.matrix(sort(h[!user_ratings_weighted[d,]],decreasing=TRUE))

# h <- melt(ratings, id=c("user_id","movie_id"), measure="ratings")
# hh <- acast(h, user_id ~ movie_id ~ variable)
## gg <- as.matrix(hh[,,1])

# user_vectors_dim <- length(idf$tags)*length(users$user_id)
# user_vectors <- data.frame(colname1=character(user_vectors_dim),
colname2=character(user_vectors_dim), colname3=numeric(user_vectors_dim))
# colnames(user_vectors) <- c("user_id","tags","tfidf")
# user_vectors$user_id <- rep(users$user_id,each=length(idf$tags))
# user_vectors$tags <- rep(idf$tags,length(users$user_id))
```

```
# ratings$movies==tfidf[1,2]

# user_movies <- as.data.frame(table(user_ratings$movie_id,user_ratings$user_id))
# colnames(user_movies) <- c("movie_id","user_id","freq")
# user_vector_fun <- function(c) (if(user_movies[c,3] !=0) user_vectors[user_vectors
$user_id==user_movies[c,2],3] <- user_vectors[user_vectors$user_id==user_movies[c,2],3] +
user_movies[c,3]*tfidf[tfidf$movie_id==user_movies[c,1],3])


# h <- mapply(user_vector_fun,1:nrow(user_movies))


# user_vectors[1,4] <- user_vectors[1,4]+y[1,3]*tfidf[tfidf$movies==y[1,1],3]
```