

Material for the Movie Recommendation task

Task Predict the user's rating for a given movie

Objects Movies and users

Target value Movie rating (1 to 5)

rating	Integer (1 to 5)	movie rating
--------	------------------	--------------

Feature vector – 32 values of different types

feature	type	definition
movie	categorical	movie id
user	categorical	user id
timestamp	numerical	movie rating timestamp
age	integer	user age
gender	categorical	user gender
occupation	categorical	user occupation
zip	categorical	user zip code
title	categorical	movie title
release_date	categorical	movie release date
genre_action	binary	movie genre action
genre_adventure	binary	movie genre adventure
genre_animation	binary	movie genre child
genre_child	binary	movie genre action
genre_comedy	binary	movie genre comedy
genre_crime	binary	movie genre crime
genre_documentary	binary	movie genre documentary
genre_drama	binary	movie genre drama
genre_fantasy	binary	movie genre fantasy
genre_filmnoir	binary	movie genre film-noir
genre_horror	binary	movie genre horror
genre_musical	binary	movie genre musical
genre_mystery	binary	movie genre mystery
genre_romance	binary	movie genre romance
genre_scifi	binary	movie genre sci-fi
genre_thriller	binary	movie genre thriller
genre_war	binary	movie genre war
genre_western	binary	movie genre western
imdb_url	categorical	movie IMDb url
imdb_rating	continuous	movie IMDb rating
directors	categorical	movie directors (a maximum of 3)
writers	categorical	movie writers (a maximum of 3)
stars	categorical	movie stars (a maximum of 3)

Use the R script load-mov-data.R to upload the MOV data into R

```
#####
###           R code for loading the MOV data           ###
###           Barbora Hladka, Martin Holub             ###
###           ESSLLI 2015                               ###
###           http://ufal.mff.cuni.cz/esslli2015           ###
#####

## get examples
> examples <- read.csv("mov.development.csv", sep="\t")
> names(examples)
 [1] "movie"          "user"          "rating"
 [4] "timestamp"     "age"          "gender"
 [7] "occupation"    "zip"          "title"
[10] "release_date"  "genre_action" "genre_adventure"
[13] "genre_animation" "genre_children" "genre_comedy"
[16] "genre_crime"    "genre_documentary" "genre_drama"
[19] "genre_fantasy" "genre_filmnoir" "genre_horror"
[22] "genre_musical" "genre_mystery" "genre_romance"
[25] "genre_scifi"   "genre_thriller" "genre_war"
[28] "genre_western" "imdb_url"      "imdb_rating"
[31] "directors"     "writers"      "stars"

## convert timestamp to date object
> examples$timestamp <- as.POSIXct(examples$timestamp, origin="1970-01-01")

## get votes
> votes <- examples[,c(2,1,3,4)]
> names(votes)
 [1] "user"          "movie"          "rating"          "timestamp"
> votes[64138,]
      user movie rating          timestamp
      43  482     4      1997-10-04 18:10:21

## get users
> u <- unique(examples[,c(2,5:8)])
> attach(u); users <- u[order(user),]; detach(u)
> names(users)
 [1] "user"          "age"          "gender"          "occupation" "zip"
> users[43,]
 user age gender occupation zip
  43  29   F   librarian  20854

## get movies
> movies <- unique(examples[,c(1,9:33)])
> names(movies)
 [1] "movie"          "title"          "release_date"
 ...
[25] "writers"        "stars"
> movies[482,]
 movie title          release_date ...          stars
 482 Some Like It Hot (1959) 01-Jan-1959 ... Marilyn Monroe,Tony Curtis,Jack Lemmon
```