

# 40.016 The Analytics Edge

## Recommendation systems (Part 2)

Stefano Galelli

SUTD

Term 6, 2020

# Outline

- Recommendation systems
- Collaborative filtering
- R implementation

# Recommendation systems

- Personalize the user experience for online applications
- Leverage data on items and customers (e.g., likes, purchase history)
- A key challenge: they must be fast and accurate
- Common underlying analytics: clustering, collaborative and content filtering

# Recommendation systems

**Collaborative filtering:** Recommendations are based on attributes of users. Each user is represented by a vector of items where the  $i$ -th entry gives the customer's rating of the  $i$ -th item. This vector will typically have many empty entries (only a small fraction of the items is ranked or purchased).

# Recommendation systems

**Content filtering:** Recommendations are made based on attributes of items. Each item is represented by a set of attributes (e.g., genre of movie, keywords, or webpage).

# Recommendation systems

**Hybrid:** This is a combination of both collaborative and content filtering.

# Collaborative filtering

## Baseline model

Predict the rating based on the item average popularity:

$$b_{ui} = \bar{r}_i$$

# Collaborative filtering

## User-based collaborative filtering

Based on:

- 1 Identifying users whose ratings are similar to those of the active user, and
- 2 Using their ratings on other items to predict what the active (current) user will like



# Collaborative filtering

## User-based collaborative filtering

To measure the similarity between users  $u$  and  $v$ , we can use the *Pearson correlation coefficient*

$$S_{uv} = \frac{\sum_{i \in I_u \cap I_v} (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in I_u \cap I_v} (r_{ui} - \bar{r}_u)^2 \sum_{i \in I_u \cap I_v} (r_{vi} - \bar{r}_v)^2}}$$

# Collaborative filtering

## User-based collaborative filtering

To predict the rating, we can use the information on a set of neighbors of user  $u$ :

$$p_{ui} = \frac{\sum_{v \in N_u} r_{vi}}{|N_u|}$$

# Collaborative filtering

## User-based collaborative filtering

A slightly more sophisticated model:

$$p_{ui} = \frac{\sum_{v \in N_u} S_{uv} r_{vi}}{\sum_{v \in N_u} S_{uv}}$$

# Collaborative filtering

## User-based collaborative filtering

A modelling trick, normalization:

$$p_{ui} = \bar{r}_u + \frac{\sum_{v \in N_u} S_{uv}(r_{vi} - \bar{r}_v)}{\sum_{v \in N_u} S_{uv}}$$

# Collaborative filtering

## Assessing performance

$$RMSE = \sqrt{\sum_i \frac{(p_{ui} - r_{ui})^2}{n}}$$

# R implementation

`ratings` **dataset**

# R implementation

## Data preparation

# R implementation

## Baseline model 1



# R implementation

## Baseline model 2

# R implementation

## User-based model

# R implementation

## User-based model

# References

- Teaching notes.