Bored of seeing the “same” recommended content everyday? Measure the diversity of them!
Our Python library provides the complete workflow for measuring diversity in recommendation systems

BOSTON UNIVERSITY

Measuring Diversity in Recommendation Systems

Nanna Katrín Hannesdóttir, Zhaoyuan Fu

Dataset Formats

movielens

Rating data sets from the MovieLens web site, in the formats of:

pandas

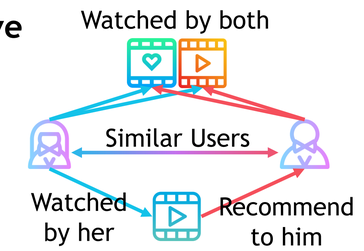
Data analysis and manipulation tool

APACHE Spark™

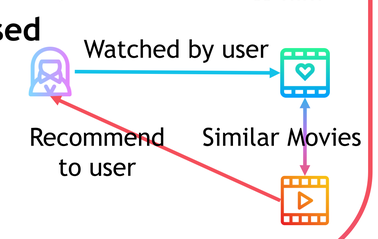
Unified engine for large-scale data analytics

Recommendation Systems

Collaborative Filtering



Content-based Filtering



Diversity Metrics

Similarity between items:

$$Cosine\ Similarity(i,j) = \frac{|M_t^{l(i,j)}|}{\sqrt{|M_t^{l(i)}|}\sqrt{|M_t^{l(j)}|}}$$

$M_t^{l(i)}$: users liked item i

Overall similarity:

$$Intra-List\ Similarity(IL) = \frac{1}{|M|}\sum_{u\in M}\frac{1}{\left(\frac{N_r(u)}{2}\right)}\sum_{i,j\in N_r(u), i<j} Cosine\ Similarity(i,j)$$

$N_r(u)$: recommendations for user

Diversity:

$$Diversity = 1 - IL$$

Measure Functions

Co-Occurrence Diversity:

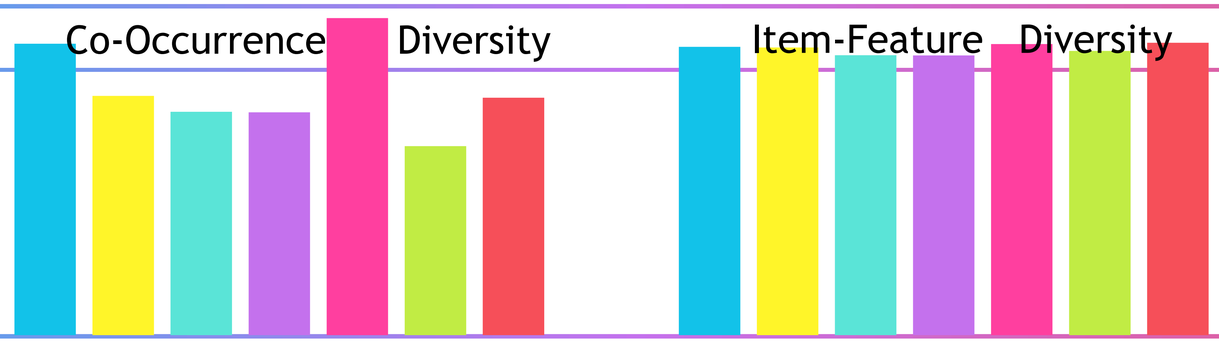
For each item pair, measure the diversity between the user groups that it is recommended to

Item-Feature Diversity:

For each user, measure the diversity between items that are recommended

Results

Alternating Least Squares
Singular Value Decomposition
Neural Collaborative Filtering
Graph Convolution Network
Restricted Boltzmann Machine
Simple Algorithm for Recommendation
Wide linear & Deep neural



80%

Winners: ALS and RBM!
For their both Co-Occurrence Diversity and Item-Feature Diversity exceeded 80%!

Boston University College of Engineering