Non-personalised collaborative filtering (CF) & user-based CF

Dom Jina

January 18, 2024

non-personalised collaborative filtering

Collaborative filter - tell me whats popular among peers

User ratings and actions are noisy measurements
Relevance is a user's perogative

Feedback model

Explicit feedback - Reviews, likes, comments, shares
If not a lot of rating data, a lot of noise i.e. 8/10 for 9 reviews > 7/10 for 10,000 reviews

cons: are ratings reliable and accurate

implicit feedback - abundant data from user actions - views, clicks, reads, buys, etc

Not direct expressions of preferences
Didn't click: bad or didn't see?
cons: lots of storage needed

collaborative filtering

Leveage "wisdom of the crowds"

Taking the results of others to predict the values of an unknown.

Problem: will be the same for all users, however not all users will like the same thing

Could try segment users, i.e. age groups, income, location, etc

Not fully personalised.

Use sessions and associtions

Use historical profiles - may introduce spurious associtions

Use transaction data - may limit follow up sales

Time constrained profiles - offer a compromise

Association rule mining

Given a set of transactions, find rules that will predict the occurence of an item based on occurence of others.

Look at frequent items sets and partition them

Rules are not from one item to another

They are from one set to another i.e. one or more than one (set)

item set - A colleciton of one or more items, a k-item set is exactly k items

support count is frequency of occurence of an itemset

support fraction of all transactions that contain an itemset

Association rule between one set and another

Rule evaluation metrics fraction of transactions that contain all items of X and Y

Confidence of an item measures how often transactions containing Y appear within the transactions that contain X