Popularity & Content based

Recommendation System

greatlearning
Learning for Life

Agenda

- Recommendation systems overview
- Types of recommendation systems
- Popularity based recommendation systems
- Content based recommendation systems
- Similarity fundamentals

greatlearning
Learning for Life

Recommendation Systems (TOC)

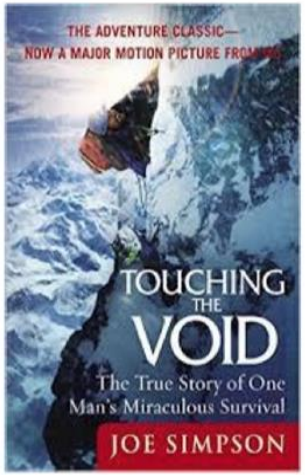
S. No.	Topic	Scope	Objective
1	Introduction to recommendation systems	Discuss what recommendation system is and why is it imp	To understand the motivation and need for recommendation systems
2	Types of recommendation systems	Discussing various types of recommendation system	To understand various methods to build a recommender
3	Popularity based recommender	Discuss theory behind popularity based model	How to recommend popular items to a new user
5	Content based model	Advantages, BOW, TF	To understand how to recommend items using content based
6	Similarity fundamentals	What and why of similarity functions, different types, their application, Euclidean, Jaccard, Pearson, Cosine	To understand the fundamental of similarity, to be able to use write method to calculate similarity distance

greatlearning
Learning for Life

Overview

- Provides most relevant item/information to a user
- Uses pattern from the existing data
- Build using popularity and similarity concepts

Recommendation systems classical example



Published in 1988

Published in 1996

In 1988, a British mountain climber named Joe Simpson wrote a book called Touching the Void, a harrowing account of near death in the Peruvian Andes. It got good reviews, only a modest success, it was soon forgotten. Then, a decade later, a strange thing happened. Jon Krakauer wrote Into Thin Air, another book about a mountain-climbing tragedy, which became a publishing sensation. Suddently, Touching the Void started to sell again."...The Long Tail by Chris Anderson

Recommendation System Applications

- Netflix recommends movies
- E-commerce sites recommends item based on our purchase history
- Youtube recommends videos
- Song recommendations by spotify
- Food recommendations by restaurants
- Social media sites recommends content
- Courses in e-learning
- Jobs
- Advertising messages

Why Recommendation Systems ?

- Increase in sales by personalizing offers
- Enhanced customer experience
- More time spent on platform
- Helps user to find item of their interest
- Helps provider to deliver an item to the right user

Does it work?

NETFLIX 2/3 Rented Movie are from recommendation

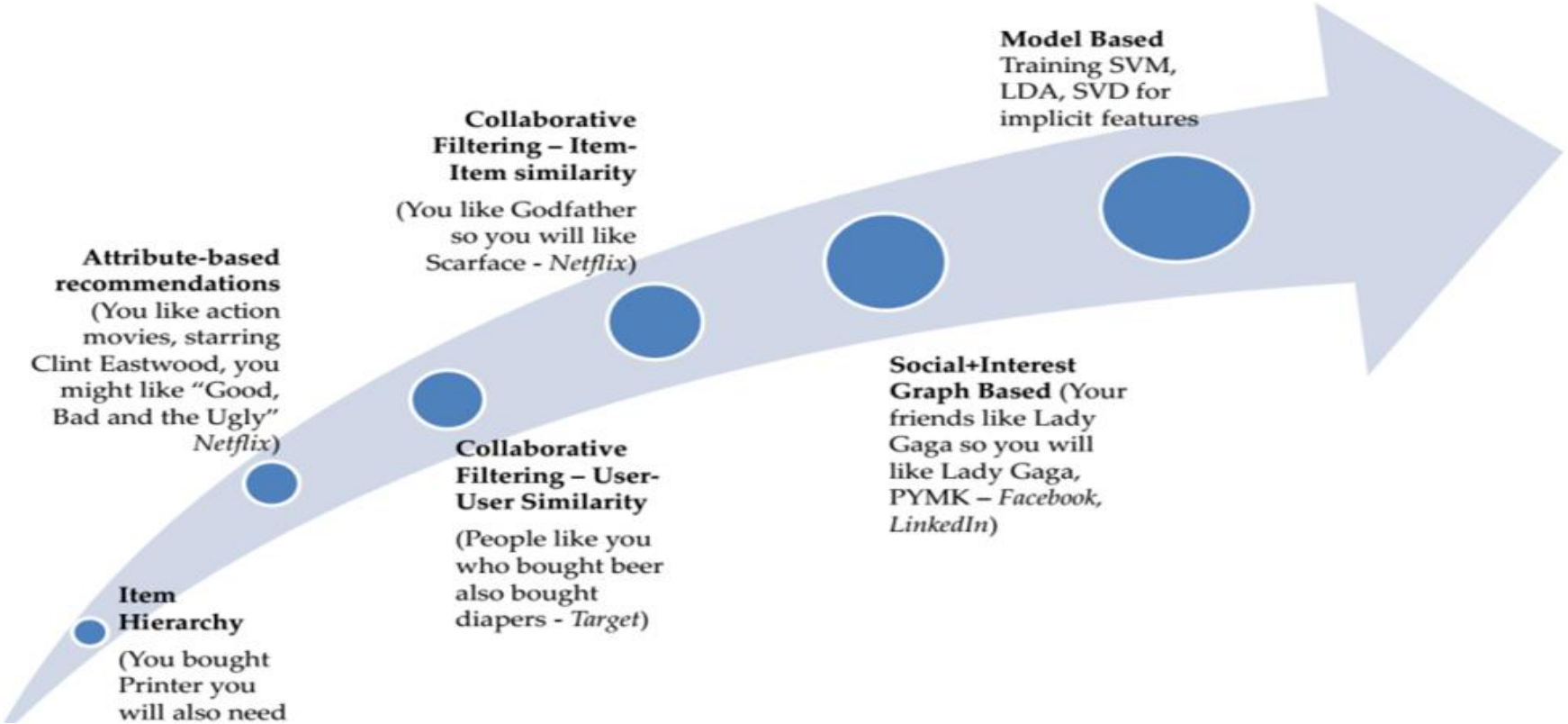
Google news 38% more Click through are due to recommendation.

amazon 35% sales are from recommendation.

Uses

1. Prediction - Estimate of rating a user would assign
2. Recommendation - List of candidate items for user to consider
3. Classification - to predict whether a user will like a product or not using features of both users and products

Evolution to present



Types of recommendation systems

- Popularity based
- Content based
- Collaborative Filtering
  - User based
  - Item based
- Matrix factorization
- Association Rule
- Hybrid

Popularity based

- Recommends items which are popular/purchased/rated by most users.
  Example: Popular news articles
- It uses:
  - Context
  - Purchase history
  - User-Item features
- Recommendations are non-personalized
- Every user gets the same recommendations irrespective of their taste and preferences.
- Works with the trend
- Does not suffer from cold start problem

Steps involved for a popularity based model

1. Read the dataset
2. Use group by or SQL merge concept to get the ratings for each item/movie
3. Take mean of all the item/movie
4. Sort it in descending order
5. Recommend top ten or five to a new user

Notes:

- Takes only mean into consideration
- User bias can be removed by deducting mean rating of a user for each item.

Hands-on

Case study

Content based model

- Recommendations are based on information on the content of items rather than on other user's opinions

Idea

- If you like an item then you will also like a **similar** item
- Recommend items to user **x** similar to previous items rated highly by **x**

- No need for the data on other users
- No cold start and sparsity Problem
- Technique that can be used- **cosine similarity**

Content based model

Advantages	Challenges
• Do not require a lot of user's data • Only item data is sufficient • Does not suffer from cold start problem • Less expensive to build and maintain	• Feature availability • Less dynamic • Diversity is missing

Similarity fundamentals

- **Euclidean distance**

$$D_{euclidean} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots\ldots (x_n - y_n)^2}$$

  - Does not efficiently use for comparisons

- **Cosine Similarity**

$$D_{cosine} = \frac{\sum_i x_i . y_i}{\|x\| \|y\|}$$

  - Similarity is the cosine of the angle between the 2 vectors.
  - Closer the vectors, smaller will be the angle and larger the cosine.

Similarity fundamentals

- **Pearson Similarity**

$$D_{pearson} = \frac{\sum_i (x_i - \bar{x}).(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

  - Similarity is the pearson coefficient between the two vectors
  - Pearson correlation is invariant to shift.
  - Pearson and cosine similarity are invariant to scaling
  - The range of similarity varies between -1 to 1.
- **Jaccard Similarity**

$$D_{jaccard} = \frac{|A \cap B|}{|A \cup B|}$$

  - used to comparing the similarity and diversity of sample set.

Rating Matrix

Texts are converted to a document term matrix using cosine similarity

Document 1- Pride of India : A glimpse into India's scientific heritage

Document 2- Makers of modern India

Document 3 - The logic of scientific discovery

Test doc- Scientific India

Recommendation order would be - D1, D2/D3

Term Frequency

D1 - Machine Learning is fun

Document	Machine	Learning	is	fun
TF	1	1	1	1
Normalized TF	0.25	0.25	0.25	0.25

- Similarly we need to create normalized TF for each document
- Calculate cosine similarity between test doc and existing docs
- Recommend based on similarity score

Implementation of a content based model

1. Read data
2. Preprocessing
3. Distance matrix calculation
4. Check similarity
5. Recommend

Hands-on

Case study

greatlearning
Learning for Life

Thank you!

Happy Learning :)