Recommender Systems

COMP3008

Lauren Ansell

Introduction

Today's topics:

- Recommender systems

Session learning outcomes - by the end of today's lecture you will be able to:

- Explain what a recommender system is.
- Explain and describe types of recommender systems.
- Understand some of the issues affecting recommender systems.

Recommender Systems

Recommender systems are a subclass of information filtering systems.

Their aim is to provide suggestions for items that are most pertinent to a particular user.

Three key components of recommender systems:

1. Candidate generations
2. Scoring systems
3. Re-ranking systems

Recommender Systems Use Cases

Netflix recommends movies you may like to watch.

Amazon recommends products you may be interested in purchasing.

Facebook recommends people you may want to contact and become friends with.

Statistics

Two-thirds of movies watched by Netflix customers are recommended movies.

35% of Amazon sales are a result of recommendations.

38% of click-through rates on Google News are recommended links.

Algorithms

Collaborative filtering: based on the idea that the best person to recommend something is someone who has similar tastes to you.

Content-based filtering: uses item features to recommend other items similar to what the user likes.

Influence scoring: a scoring system that categorises mentions by importance and popularity.

Collaborative Filtering

Collaborative filtering can be in two senses: a narrow one and a more general one.

In the narrower sense, collaborative filtering is a method of making automatic predictions about the interests of a user by collecting preferences or taste information from many users.

In the more general sense, collaborative filtering is the process of filtering for information or patterns using techniques involving collaboration among multiple agents, viewpoints, data sources, etc.

Typical Workflow For Collaborative Filitering

1: A user expresses his or her preferences by rating items of the system.

2: The system matches this user's ratings against other users' and finds the people with most "similar" tastes.

3: With similar users, the system recommends items that the similar users have rated highly but not yet being rated by this user.

Methodology

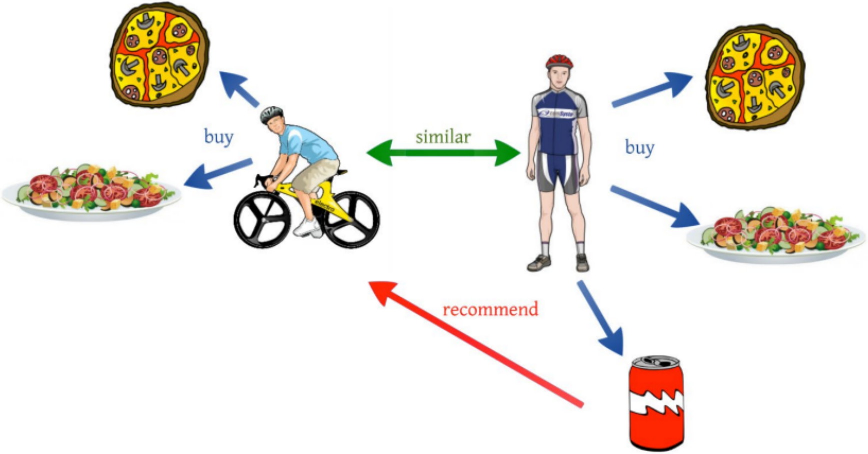
Collaborative filtering systems have many forms, but many common systems can be reduced to two steps:

1. Look for users who share the same rating patterns with the active user.
2. Use the ratings from those like-minded users found in step 1 to calculate a prediction for the active user.

Alternatively, item-based collaborative filtering, proceeds in an item-centric manner:

1. Build an item-item matrix determining relationships between pairs of items.
2. Infer the tastes of the current user by examining the matrix and matching that user's data.

Example



Types of Collaborative Filtering

Memory-based: uses user rating data to compute the similarity between users or items.

Model-based: use data mining and machine learning algorithms to predict user's rating of unrated items.

Hybird: a combination of memory-based and model-based algorithms.

Deep-learning: some generalize traditional matrix factorization algorithms via a non-linear neural architecture or leverage new model types like Variational Autoencoders.

Problems and Challenges

Data sparsity

Scalability

Synonyms

Grey sheep

Shilling attackes

Diversity and the long tail

Content-based Filtering

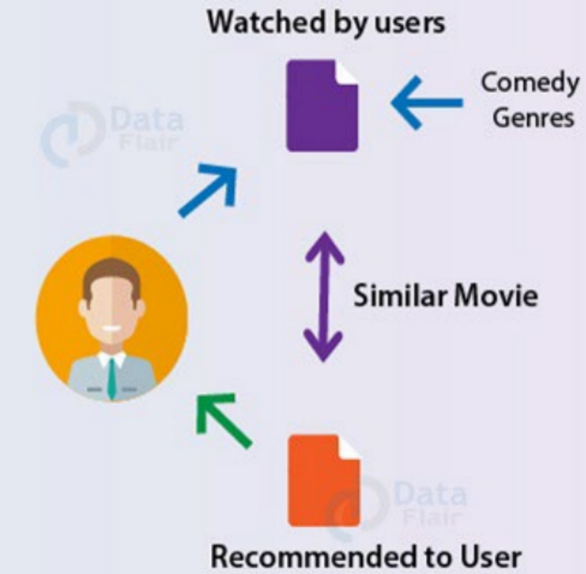
Content-based filtering methods are based on a description of the item and a profile of the user's preferences.

In this system, keywords are used to describe the items, and a user profile is built to indicate the type of item this user likes.

To create a user profile, the system mostly focuses on two types of information:

1. A model of the user's preference.
2. A history of the user's interaction with the recommender system.

Example



Watched by users
Comedy
Genres
Similar Movie
Recommended to User

A key issue with content-based filtering is whether the system can learn user preferences from users' actions regarding one content source and use them across other content types.

For example, recommending news articles based on news browsing is useful.

Content-based recommender systems can also include opinion-based recommender systems.

Hybrid Recommendation Approaches

Most recommender systems now use a hybrid approach, combining collaborative filtering, content-based filtering, and other approaches.

Several studies that empirically compared the performance of the hybrid with the pure collaborative and content-based methods and demonstrated that the hybrid methods can provide more accurate recommendations than pure approaches.

These methods can also be used to overcome some of the common problems in recommender systems such as cold start and the sparsity problem, as well as the knowledge engineering bottleneck in knowledge-based approaches.

Weighted

Switching

Mixed

Feature Combination

Feature Augmentation

Cascade

Meta-level

Example: Twitter's Recommendation Algorithm

The algorithm aims to filter down from roughly 500 million Tweets to the ones you see on the **For You** timeline.

The Twitter recommendation algorithm is constructed from several interconnected services and jobs.

Twitter also recommender systems for:

- Search
- Explore
- Ads

Evaluating Recommender Systems

Performance measures

Evaluation is important in assessing the effectiveness of recommendation algorithms.

The commonly used metrics are the mean squared error and root mean squared error.

Evaluating the performance of a recommendation algorithm on a fixed test dataset will always be extremely challenging as it is impossible to accurately predict the reactions of real users to the recommendations.

Hence any metric that computes the effectiveness of an algorithm in offline data will be imprecise.

Beyond Accuracy

Diversity

Recommender persistence

Privacy

User demographics

Robustness

Serendipity

Trust

Labelling

Reproducibility

The topic of reproducibility seems to be a recurrent issue in some Machine Learning publication venues, but does not have a considerable effect beyond the world of scientific publication.

More recent work on benchmarking a set of the same methods came to qualitatively very different results whereby neural methods were found to be among the best performing methods.

Operators of recommender systems find little guidance in the current research for answering the question, which recommendation approaches to use in a recommender systems.

Summary

Recommender systems

- Computer systems that make recommendations are called recommender systems.
- Algorithms: Collaborative filtering, influence scoring, and content-based filtering.
- Collaborative filtering and content-based filter are often used together in a hybrid approaches.