

# RECOMMENDER SYSTEMS

Sarith Divakar M

Department of CSE



**L.B.S College of Engineering**

(Managed by LBS Centre for Science & Technology - A Govt. of Kerala Undertaking)

Online Session: 07-04-2020

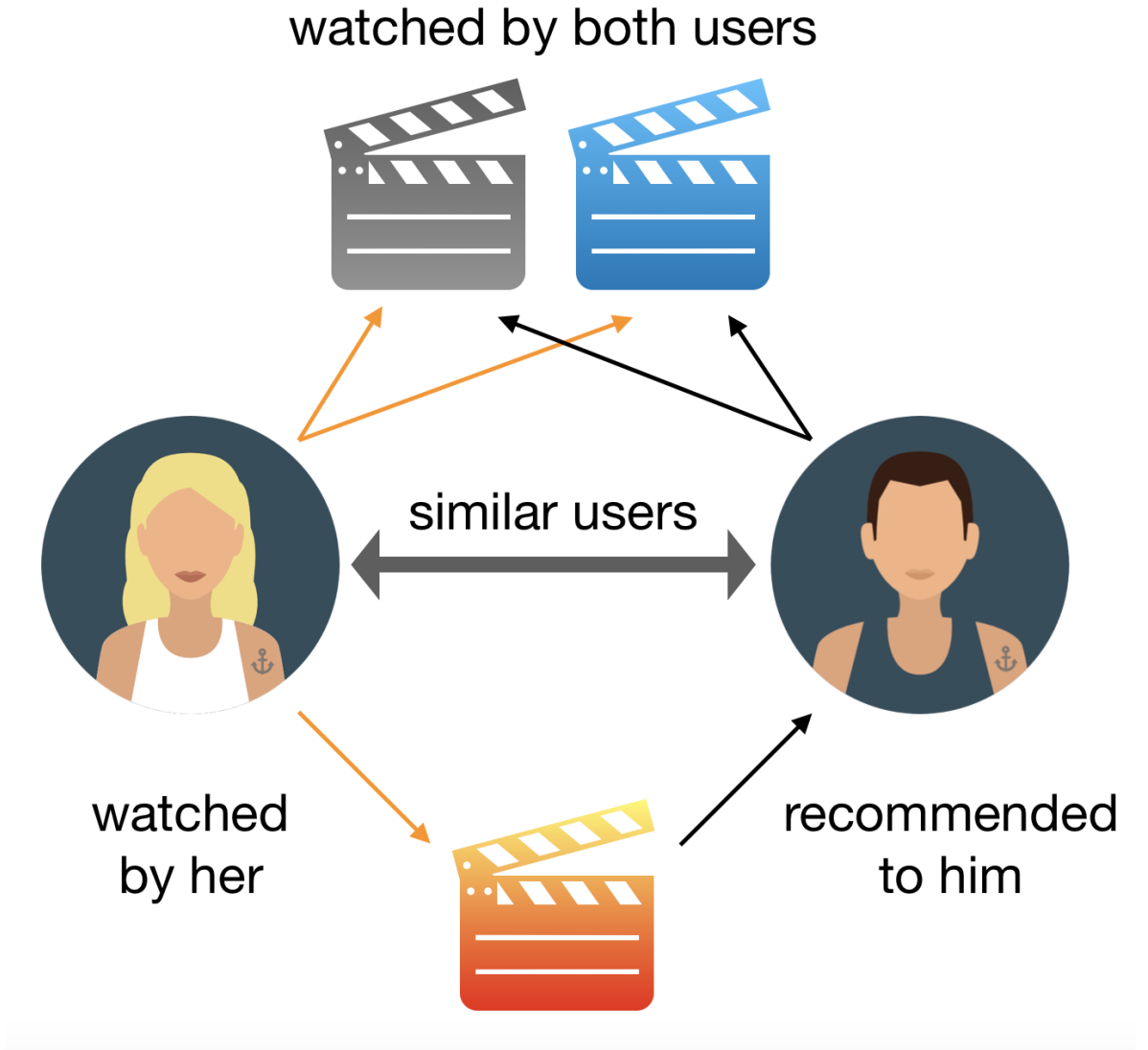
# Online Videos



# Online Shopping



# Content-Based Filtering



# Content-Based Filtering

- Recommend items based on two information sources:
  - Features of products and ratings given by users.
- Structured data: Each item consists of the same attributes and the possible values for these attributes are known.
- Unstructured: Such as text, different techniques have to be used in order to learn the user profiles
  - Free text can then be translated into more structured data by using a selection of free text terms as attributes.
  - Techniques like TF-IDF (term frequency/inverse document frequency) can then be used to assign weights to the different terms of an item.
- Semistructured: Convert the text into structured data

# Machine learning techniques used

- Classification algorithm is invoked for each user based on his or her ratings on items and their attributes.
  - logistic regression, neural networks, decision trees, association rules, and Bayesian networks
- When explicit ratings are considered, the ratings are directly used for the classification task, whereas implicit ratings can be obtained using the item–user interactions
- Nearest neighbor methods can also be used to determine the labeled items that are most similar to a new unlabeled item in order to label this new item based on the labels of the nearest neighbors
- Linear classifiers, support vector machines, Naive Bayes

# Advantages

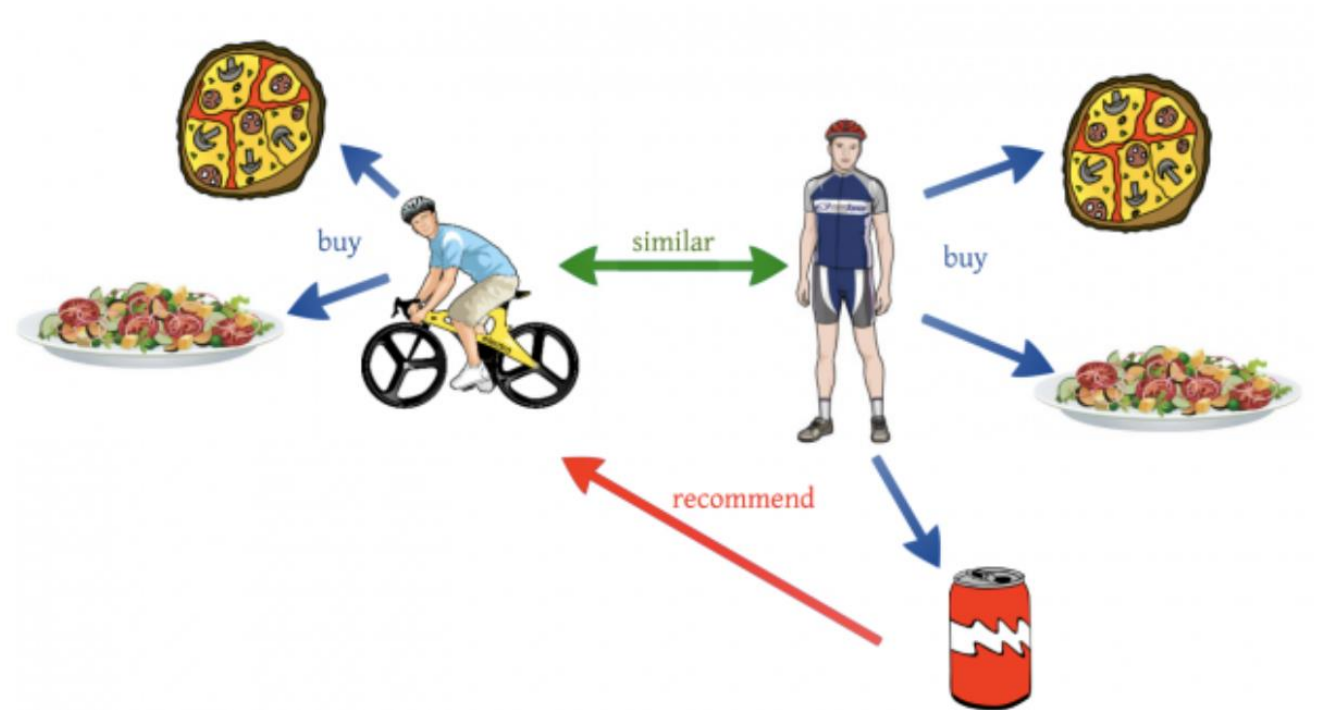
- New items (which have not received ratings before) can be recommended
- Items can also be recommended to users that have unique preferences
- Possibility to give an explanation to the user about his or her recommendations
- Only ratings of the active user are used in order to build the profile

# Disadvantages

- Only suitable if the right data are available
- The cold start problem for new users: old ratings potentially influence the recommendation too much
- Over-specialization: techniques focus on items similar to the previously bought items.



# Collaborative Filtering



# Collaborative Filtering

- Collaborative filtering, also called *social filtering*
- The main idea is to recommend items based on the opinions of other users.
- Two types:
  - user-based collaborative filtering
  - item-based collaborative filtering
- One way to calculate similarity between users or items is to use a user-item matrix that contains information on which user bought what item.

# User-based collaborative filtering

- Items will be recommended to a user based on how **similar users** rated these items

# Item-based collaborative filtering

- Items will be recommended to a user based on how this user rated **similar items**.

# Ratings

- A scalar rating can be a number or an ordinal rating.
- A binary rating consists of two possibilities, such as good or bad.
- Finally, unary ratings indicate that a user has had an interaction with an item, such as a click on an item or a purchase
- Explicit ratings can be obtained by requesting a user to rate a certain item.
- Implicit ratings are obtained by associating a rating with a certain action, such as buying an item

# Neighbourhood-based algorithms

- Similarity measure is used to calculate similarity between users (in case of a user-based algorithm) or items (in case of an item-based algorithm).
- Subset of users or items is selected that functions as the neighborhood of the active user or item
- Algorithm predicts a rating based on the active user's or item's neighborhood, typically giving the highest weight to the most similar neighbors

# Advantages

- Collaborative filtering does not restrict the type of items to be recommended.
- It manages to deliver recommendations to a user even when it is difficult to find out which specific feature of the item makes it interesting to the user or when there is no easy way to extract such a feature automatically
- Collaborative filtering is believed to recommend more unexpected items (that are equally valuable) than content-based techniques

# Disadvantages

- Sparse data can be a problem for such a technique
- *Cold start problem*: New items cannot easily be recommended because they have not been rated yet; therefore, new users cannot easily receive recommendations because they have not yet rated items
- Privacy could also be a problem because collaborative filtering needs data on users to give recommendations or could generate trust issues because a user cannot question the recommendation.



# Demographic Filtering

- Based on demographic information of the user
- The main challenge is to obtain the data
  - Explicitly done by asking for information from users
  - Analytical techniques could be used to extract information linked to the interactions of the users with the system
- Advantage
  - There is not always a need for a history of user ratings of the type that is required in collaborative and content-based approaches
  - Segments can be used in combination with user–item interactions in order to obtain a high-level recommender system
- Disadvantage
  - Cold start problem for new users and new items, as well as the difficulty in capturing the data, which is highly dependent on the participation of the users.

# Knowledge-Based Filtering

- The main difference with regard to the other techniques resides in the data sources used
- Additional inputs consisting of constraints or requirements are provided to the recommender system typically by allowing a dialog between the user and the system
- Two categories
- Constraint based recommenders: Systems meeting a set of constraints imposed by both users and the item domain.
- Case-based recommenders: The goal is to find the item that is most similar to the ones the user requires

# Advantage

- Can be used when there is only limited information about the user, hence avoiding the cold start problem
- Expert knowledge is used in the recommender system
- Constraint-based recommender system can help customers actively, for example, by explaining products or suggesting changes in case no recommendation is possible

# Disadvantage

- Require some effort concerning knowledge acquisition, knowledge engineering, and development of the user interface
- It can be difficult when the user is asked to provide the system with an example if the number of items in the recommendation system is very high.
- It may be difficult or impossible for the user to provide an example that fits the user's needs.

# Types of Hybrid Filtering

1. Weighted: recommendation scores of several recommenders are combined by applying specific weights
2. Switching: recommendations are taken from one recommender at a time, but not always the same one.
3. Mixed: recommendations for multiple recommenders are shown to the user
4. Feature combination: Different knowledge sources are used to obtain features

5. Feature augmentation: A first recommender computes the features while the next recommender computes the remainder of the recommendation.
6. Cascade: Each recommender is assigned a certain priority and if high priority recommenders produce a different score, the lower priority recommenders are decisive.
7. Meta-level hybrid recommender system: First recommender that gives a model as output that is used as input by the next recommender.

# Evaluation of Recommender Systems

- Two categories of evaluation metrics are generally considered:
- The goodness or badness of the output presented by a recommender system
  - Recall, Precision, F1 Score
- Time and space requirements.
  - Response time, main memory requirement, secondary storage requirement