Content-Based Recommender Systems (CBR)

Vani Koduru Shankara Murthy
Agenda

- Introduction
- Definition of CBR
- Architecture
- Advantages and Disadvantages
- Future work
What is a recommender system?

- Web is huge and growing day by day
- Complex and time consuming to find an item of interest
- Helps in making the decisions about the item of interest based on preferences given to an item by the user
- Most often used in e-commerce, movie, music and app recommendations
- Example- web browsers, based on previous browsing history and bookmarks
- Three main filtering techniques adopted are
  - Content based Systems
  - Collaborative Systems
  - Hybrid Systems
What is CBR?

• Matches the attributes of the user profile against the attributes of a content object.
• Items are represented as set of structured data/features
  Movie; actor, actress, director, year of release, rating, etc...
• **Keyword-based Retrieval Models (VSM)**
  • Parsing, stop words removal and stemming
  • tf-idf weighting, contents are represented as TDM
  • Basic string matching technique
• **Semantic Analysis**
  To address the problem of polysemy and Synonymy
  Interpret natural language documents
• **Building the user profile**
  • Monitoring and analyzing the user activity
  • Browsing history serves as the training data for machine learning algorithms
  • Probabilistic Models
• CBR uses **Cosine Similarity** measure to find contents that matches user profile
High-level architecture

Source: State of the Art and Trends by Pasquale Lops, Marco de Gemmis, Giovanni Semeraro.
Content Analyzer
- Extracts the features (keywords, n-grams) from the source
- Conversion from unstructured to structured item (item vector)
- Data stored in the repository **Represented Items**

Profile Learner
- To build user profile
- Updates the profile using the data in **Feedback** repository

Filtering Component
- Matching the user profile with the actual item to be recommended
- Uses different Strategies
  - Cosine similarity
Advantages

* **User Independence:** Recommends only the items that interest the user
* **Transparency:** Recommendation is based on the item features, explicitly list the contents features
* **New Item:** Helps in recommending new items that are not yet rated by other users

Disadvantages

* **Limited Content Analysis:** They have the limit in the number and type of the feature associated with an item
* **Overspecialization:** Recommends those items that score high with the user profile
* **Cold Start Problem:** For a new user, systems don’t have historical information to recommend items
Challenges of feeding user with surprising and more reliable data

**Collaborative Filtering**
- Recommendations based on the ratings and preferences given by other users
- Cold Start Problem – requires initial data set

**Hybrid Filtering**
- Combines the approaches of Content based and Collaborative Filtering
- Good Strategy
- Amazon
References


Cold start problem
Thank you!