Introduction to Recommendation System

Tulika Sahu
Software Developer
IBM Cloud and Cognitive Software
IBM USA

Presenter Bio

- Tulika Sahu is Master of Science from Illinois Institute of Technology, Chicago. She is presently serving as a software developer in IBM Hybrid cloud under IBM Data and AI in the Bay Area.
- Tulika is a technology enthusiast and is really passionate to contribute and help more people from varied backgrounds in Bihar to make use of latest technological advancements in the field of Data, ML and AI in their respective fields.

Industry Application

- Use Data to boost sales by predicting user's interests and recommend items that quite likely are of their interest, based on history.
- Industries that use Recommendation sys presently:

✓ Retail/E-Commerce: single item, a bundle of items, etc

✓ Movies: each movie, director, actor, movie genre, etc

✓ Music: single track, album, artist, playlist, etc

✓ Travel: tour, flight, hotel, car rental, travel package, etc

✓ Social networks: tweets, user accounts, groups, etc

Advantages

- Personalized offers and an enhanced customer experience.
- Speed up searches and make it easier for users to access content they’re interested in, and surprise them with offers they would have never searched for.
- Gain and retain customers by approaching customers who are most likely to be interested.
- Customers feel they are understood and is more likely to buy additional products.
- Company gains competitive advantage and reduced threat of losing customers to a competitors.

How it works

- User Preference review formats:

Ratings Binary Feedback

Explicit

Reviews Behaviors

Implicit

Traditional Recommendation Algorithms

- Content-Based Recommendation Algorithms
  The user will be recommended items similar to the ones the user preferred in the past, e.g., ecommerce, book/movie recsys

- Collaborative Filtering Based Recommendation Algorithms
  The user will be recommended items that people with similar tastes and preferences liked in the past, e.g., movie recsys

- Hybrid Recommendation Algorithms
  Combine content-based and collaborative filtering based algorithms to produce item recommendations.

New Types of Recommender Systems

- Context-Aware Recommender Systems
- Multi-Criteria Recommender Systems
- Group Recommender Systems
- Cross-Domain Recommender Systems
- Human Factor Based Recommender Systems
- Health Recommender Systems
- Multi-Stakeholder Recommender Systems
- Social Recommender Systems

Recommendation task

- There are 2 recommendation tasks in the area of recommendation systems:

✓ Rating Prediction:

Given a user and an item, predict Rating (u, t)

✓ Top-N Recommendations:

Given a user, provide a list of top-N item recommendations

Recommendation task as supervised learning

Recommendation problem is a process of supervised learning too

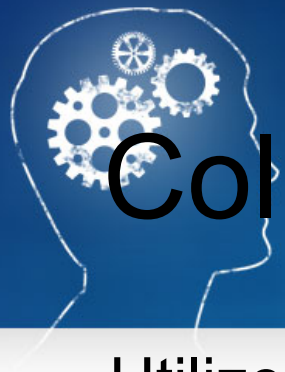
- ✓ We have a rating data
- ✓ We can split the data to training and testing set
- ✓ The truth are the real-ratings
- ✓ The predictions are the predicted ratings
- ✓ We can evaluate the strength of our predictions

Content-based system

- Make recommendations using a user's item and profile features.
- Hypothesis: If a user was interested in an item in the past, they will once again be interested in it in the future.
- Similar items are grouped based on their features.
- Historical interactions or surveys help build user profiles.

Challenges of Content-based

- Makes obvious recommendations because of excessive specialization (user A is only interested in categories B, C, and D, and the system is not able to recommend items outside those categories, even though they could be interesting to them).
- Cold start - new users lack a defined profile unless they are explicitly asked for information.



Collaborative filtering systems

- Utilizes user interactions to filter for items of interest.
- Based on assumption that if a user likes item A and another user likes the same item A as well as another item, item B; then first user maybe interested in the item B.
- We can visualize the set of interactions with a matrix, where each entry (i, j)(*i*,*j*)represents the interaction between user i*i* and item j*j*. An interesting way of looking at collaborative filtering is to think of it as a generalization of **classification** and **regression**.

Collaborative filtering Algorithms

- Memory-Based algorithms:

✓ Load rating data into memory
✓ May not have learning process
✓ Not flexible for online updates
✓ Example: Neighborhood-Based CF

- Model-Based algorithms:

✓ Usually, they are learning-based models,
✓ More effective and efficient for larger number of items to a larger number of users
✓ Example: Matrix Factorization

Challenges of collaborative filtering systems

- Cold start:

Needs enough information (user-item interactions) for the system to work. If we setup a new retail business, we cannot give recommendations until users have interacted with a significant number of items.

- Adding new users/items to the system:

whether it is a new user or item, we have no prior information about them since they don’t have existing interactions.

Evaluating Recommender Sys

- For rating prediction (error measures):
- ✓ **MAE** (Mean Absolute Error)
- ✓ **MSE** (Mean Squared Error)
- ✓ **RMSE** (Root mean Squared Error)
- Evaluating lists of recommendation (based on relevancy levels):
- ✓ Precision
- ✓ Recall
- ✓ MAP: Mean Average Precision
- ✓ nDCG: normalized Discounted Cummulative Gain
- Diversity:
- ✓ Intra-list Similarity
- ✓ Lathia's Diversity
- Implicit Feedback:
- ✓ Mean Percentage Ranking
- ✓ User-Centric Evaluation Frameworks

Conclusion

- Have a basic recommender system for a small set of users, and invest in more powerful techniques once the user base grows.
- **Data** indispensable resource. Proper storing and management of data is most important.
- Business goals will dictate the type of recommender system to implement; whether it is generating more engagement for already active customers, or pushing those infrequent customers to become more active.
- Besides defining the business goals, one should analyze and understand the information generated and thus knowledge.