Hybrid recommender Systems with LightFM

CSCI E-82 Tool Topic
Fall 2024
Sarah Pierro and Yuvraj Puri

Outline

1. Recommender systems overview
2. Cold-start problem
3. Types of recommender systems
4. Production-level
5. LightFM
6. Million Song hybrid recommender system workflow
7. Experience with LightFM
8. Example

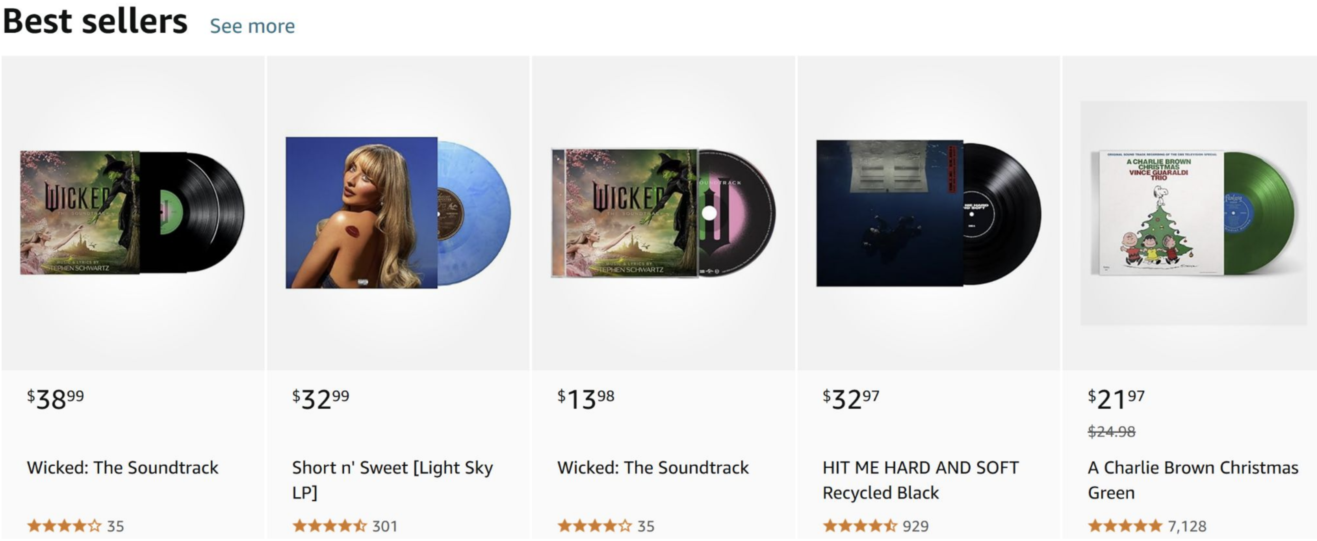
Recommendation systems: what are they?

- Basic premise: “What option is best suited for me / end user?”
- Uses: shopping, browsing, consuming media (TV, movies, YouTube)
- Considerations: “cold start problem”

“Cold start”

Users:
There is no user history for a new user. The system doesn’t know preferences of the user to make recommendations with.

Items:
Lack of interaction data for a new product. Without it, how do you know who to recommend it to or when to recommend it?

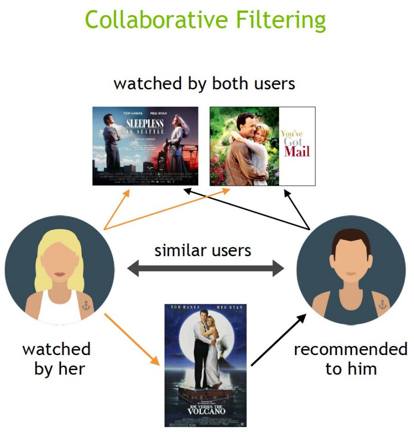


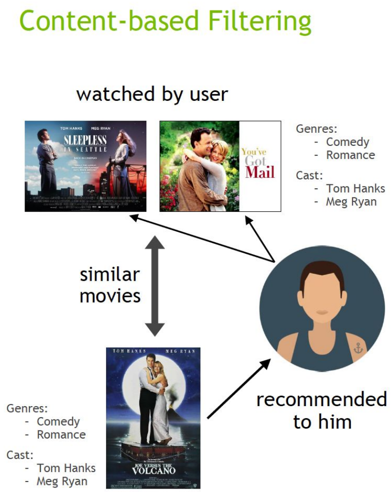
Strategies to deal with cold start: (User, Item)

- <u>Use hybrid recommender</u>
- “Preference elicitation” / onboarding: questionnaire
- Demographic-based recommendations

- Leveraging metadata (genre, tags)
- Popularity-based recommendations: recommend to users based on well-received items similar to new item

Types of recommender systems





- **Collaborative filtering**: Uses interaction data to learn patterns and make recommendations → "users who like item X also like item Y"
- **Content-based filtering**: Relies on metadata/features to
- **Hybrid recommender approaches**: combine above approaches for more comprehensive recommendation algorithm
- **Deep neural network approaches**: VAE-CF, Wide & Deep, DLRM,

- More in depth, subsets of Collaborative filtering:
  - Model based filtering
  - Memory / Session based filtering

AND PLENTY MORE!

Production level recommender systems (basic means)

- Some means they use include:
  - Candidate generation
  - Explicit feedback
  - Implicit feedback
  - Ranking

- Examples: how do they use these?
  - Netflix

LightFM

light fm

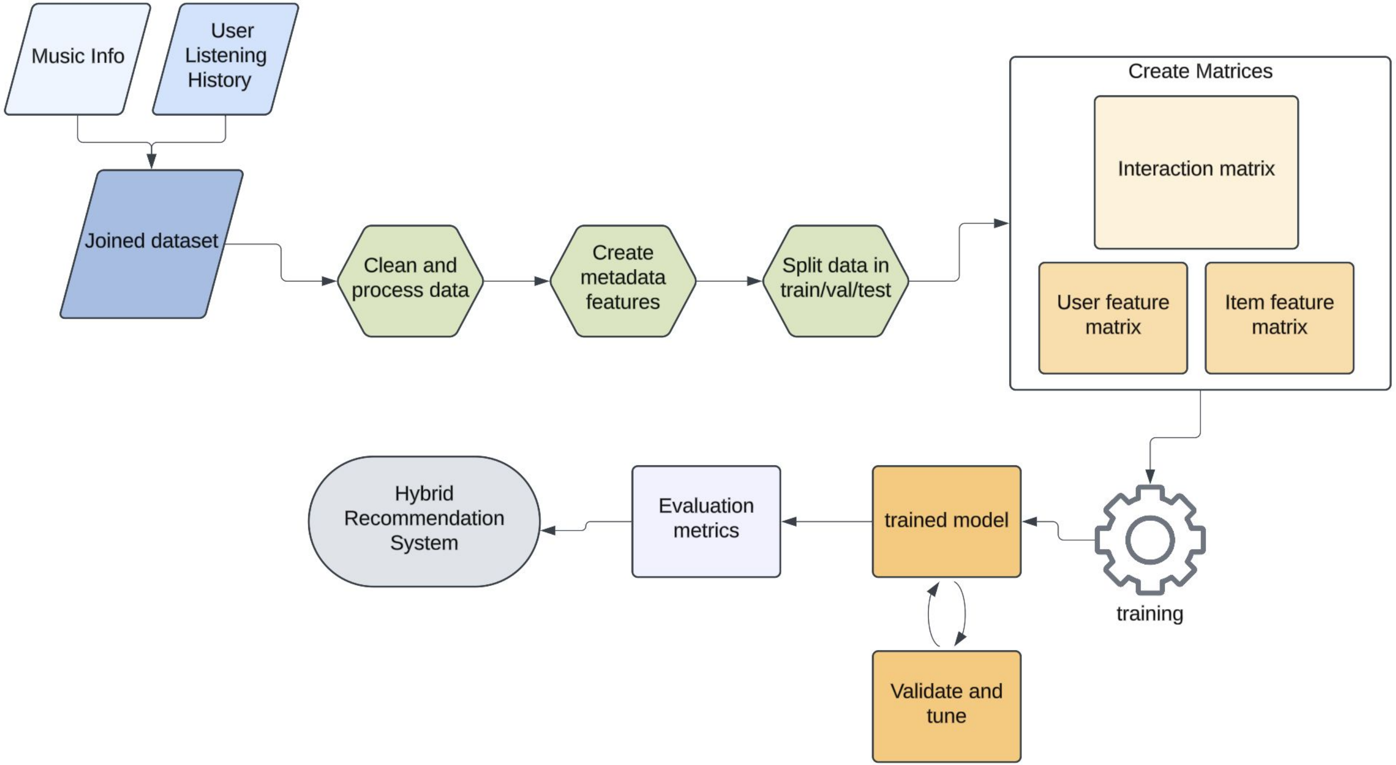
https://github.com/lyst/lightfm

- Library to implement recommendation algorithms
- **Key features:**
  - Makes building hybrid recommendation systems convenient
    - Hybrid matrix factorization model
    - User interactions and user/item metadata
  - Can model implicit or explicit feedback
  - Able to incorporate features on items and users
  - Efficient implementation of Weighted Approximate-Rank Pairwise (WARP) and Bayesian Personalized Ranking (BPR) loss
- **How it works:**
  - Maps users and items into a shared latent space
  - Choice of implicit (e.g., clicks or views) or explicit feedback (e.g., ratings or play counts)
  - Choice of loss function
  - Uses features of items and users to improve recommendations
- Performs well on cold-start recommendations
  - Existing users: use user-item interactions to make predictions
  - New users: use features about user to make predictions

LightFM on Million Song + Spotify + LastFM dataset

Data: Music streaming service with user listening history and song metadata

Goal: Recommend top K songs based on previous listens (interactions), where available, and additional features



Experience with LightFM

- Computational speed/Scalability
  - Held true to being scalable - fast training on large data (>1M rows), and seemed to handle as much data as Colab allowed
- Ease of use
  - Minimal lines of code required to initialize and fit model
- Data Preparation
  - Hardest part is data cleaning and preparation
  - Engineering additional features for user and item metadata worthwhile time investment
- Model performance
  - Able to achieve good AUC
  - Precision and recall more difficult to optimize, balancing number of top recommendations and track sample size
    - Likely gets easier with experience :-)

Example

User

Listens to rock, electronic, and pop music

	playcount
genre	
Electronic	42.0
Pop	7.0
Rock	17.0

User preferences

Top 5 recommendations

name	artist	genre	danceability	loudness	tempo	energy
You Could Be Happy	Snow Patrol	Rock	0.661224	0.790469	0.504124	0.320986
North American Scum	LCD Soundsystem	Electronic	0.900000	0.871531	0.627247	0.841997
You'll Find A Way	Santigold	Rock	0.612245	0.855457	0.844256	0.742995
Nostrand	Ratatat	Electronic	0.724490	0.811241	0.772256	0.374987
El camino	Callaghan	Rock	0.588776	0.858223	0.695180	0.920998

Resources

LightFM docs

Metadata Embeddings for User and Item Cold-start Recommendations

Million Song + Spotify + LastFM Kaggle dataset

Million Song Dataset (additional info)

Hybrid Recommender Systems with LightFM Medium Article

https://medium.com/@zaiinn440/one-stop-guide-for-production-recommendation-systems-9491f68d92e3

https://www.nvidia.com/en-us/glossary/recommendation-system/

https://medium.com/@markmilankovich/the-cold-start-problem-for-recommender-systems-89a76505a7

Thank you!