

EECS 598-008: Special Topics, Winter 2019
Advanced Data Mining

Movie Recommender System

-A Hybrid Neural Recommender

Speaker: Weijie Sun, Yuanhang Cui, Shiyu Wang

Problem

- ❖ **Users: Get information interested in**
- ❖ **Content Provider: Make contents stand out from the crowd**
- ❖ **Predict “ratings” a user would give to an item or “interactions” between them**

Challenges

How do we win the cold-start problem?

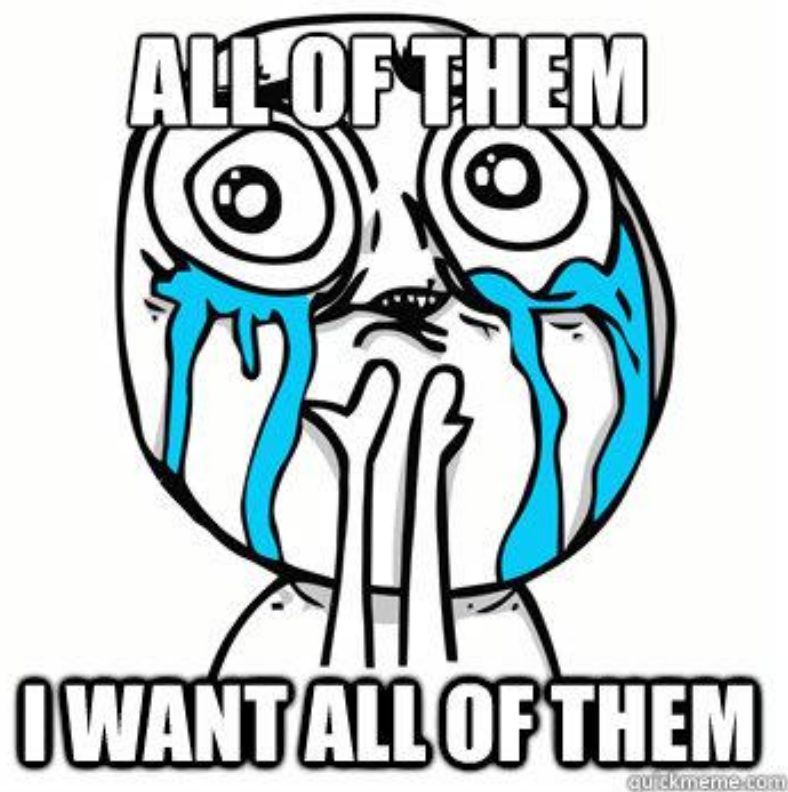
How do we improve the recommendation quality?

How do we make recommendations based on multiple factors?

Famous Recommendation Approaches

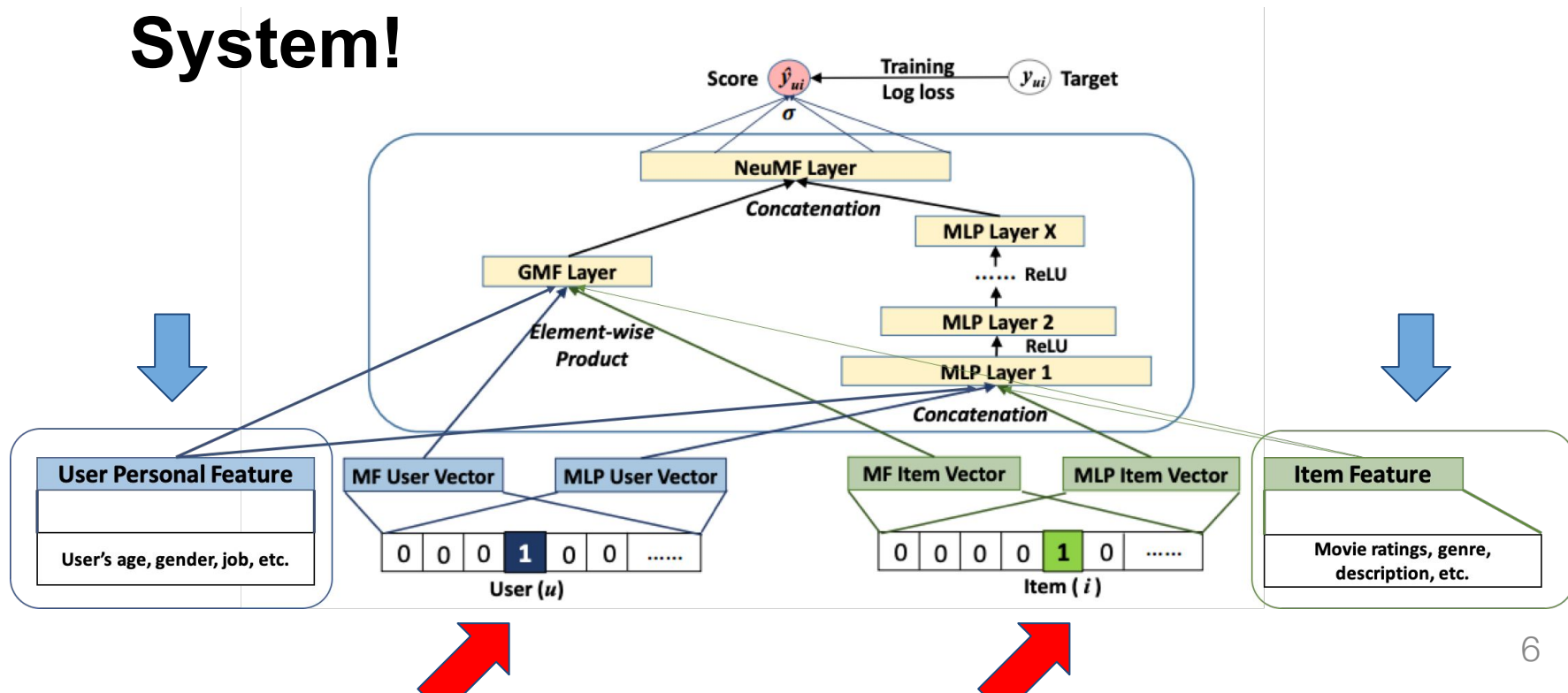
Recommendation algorithms	Advantages	Disadvantages
Content based	<ol style="list-style-type: none">1. Recommendation result is intuitive and easy to interpret;2. No need for users? Access history data;3. No new item problem and no sparsity problem;4. Supported by the mature technology of classification learning.	<ol style="list-style-type: none">1. Limited by the features extraction methods;2. New user problem;3. The training of classifier needs massive data;4. Poor scalability.
Collaborative filtering	<ol style="list-style-type: none">1. No need for professional knowledge;2. Performance improving as the increasing of the user number;3. Automatic;4. Easy to find user's new interesting point;5. Complex unstructured item can be processed. eg. Music, Video, etc.	<ol style="list-style-type: none">1. Sparsity problem;2. Poor scalability;3. New user and new item problem;4. The recommendation quality limited by the history data set

Which one should we use?



Proposed Approach

❖ CB-CF Hybrid Neural Recommender System!



Data

- ❖ **MovieLens**
 - **20 million ratings**
 - **465,000 tag applications**
 - **27,000 movie**
 - **138,000 users**

tagId	genome_tag	movieId	relevance	title	genres	userId	tag	tags_timestamp	rating	rating_timestamp
1	7	1	0.025	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	1644	Watched	1417736680	3.5	1421135243
2	007 (series)	1	0.025	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	1741	computer animatio	1183903155	4	1038421883
3	18th century	1	0.05775	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	1741	Disney animated f	1183933307	4	1038421883
4	1920s	1	0.09675	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	1741	Pixar animation	1183934770	4	1038421883
5	1930s	1	0.14675	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	1741	TrfB@a Leoni does	1245093573	4	1038421883
6	1950s	1	0.217	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	2299	Pixar	1403306852	3	1403306845
7	1960s	1	0.067	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	3596	animation	1290312028	1	1290312052
8	1970s	1	0.26275	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	3596	family	1290312032	1	1290312052
9	1980s	1	0.262	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	3596	Tom Hanks	1290312036	1	1290312052
10	19th century	1	0.032	Toy Story (1995)	Adventure Animation Children Comedy Fantasy	6233	Pixar	1387079572	4.5	1390443274

Evaluation Methodology

- ❖ Hit Ratio (HR) on top-K list

$$HR@K = \frac{NumberofHits@K}{|GroundTruth|}$$

- ❖ Normalized Discounted Cumulative Gain (NDCG) on top-K list

graded relevance at position i

$$NDCG@k = Z_K \sum_{i=1}^K \frac{2^{r_i} - 1}{\log_2(i + 1)}$$

Normalizer