Heterogeneous Knowledge in Recommendation Systems

Mentored by - Dr. Muneendra Ojha

By Aditi Agrahari, Jai Kumar Dewani and Utkarsh Raj Singh

Objective

- Study of prevalent methods of recommendation
- Integrating heterogeneous knowledge from multiple sources.
- Using knowledge graph to generate additional input features for the neural network in the hope of improving the recommendation system.

Recommender Systems

- These are algorithms aimed at suggesting relevant items to users.
- Items can be movies to watch, text to read, products to buy or anything else depending on industries.

Mainly of two types:

1. Content based filtering
2. Collaborative based filtering

Content Based Filtering

It uses similarity between items to recommend items, like items similar to what the user likes, dislikes and based on the user's previous actions.

Advantage:

- **The model predictions are not dependent upon other user data and hence overcomes user cold start**

Disadvantage:

- **Since the features are hand-engineered to some extent, this technique requires a lot of domain knowledge.**

Collaborative filtering

To address some of the limitations of content-based filtering,

- Collaborative filtering uses similarities between users and items simultaneously to provide recommendations.
- This allows to recommend an item to user A based on the interests of a similar user B.
- This requires embedding that can be learned automatically.

Mainly of two types

- Matrix Factorization based method
- Deep Neural Network based method

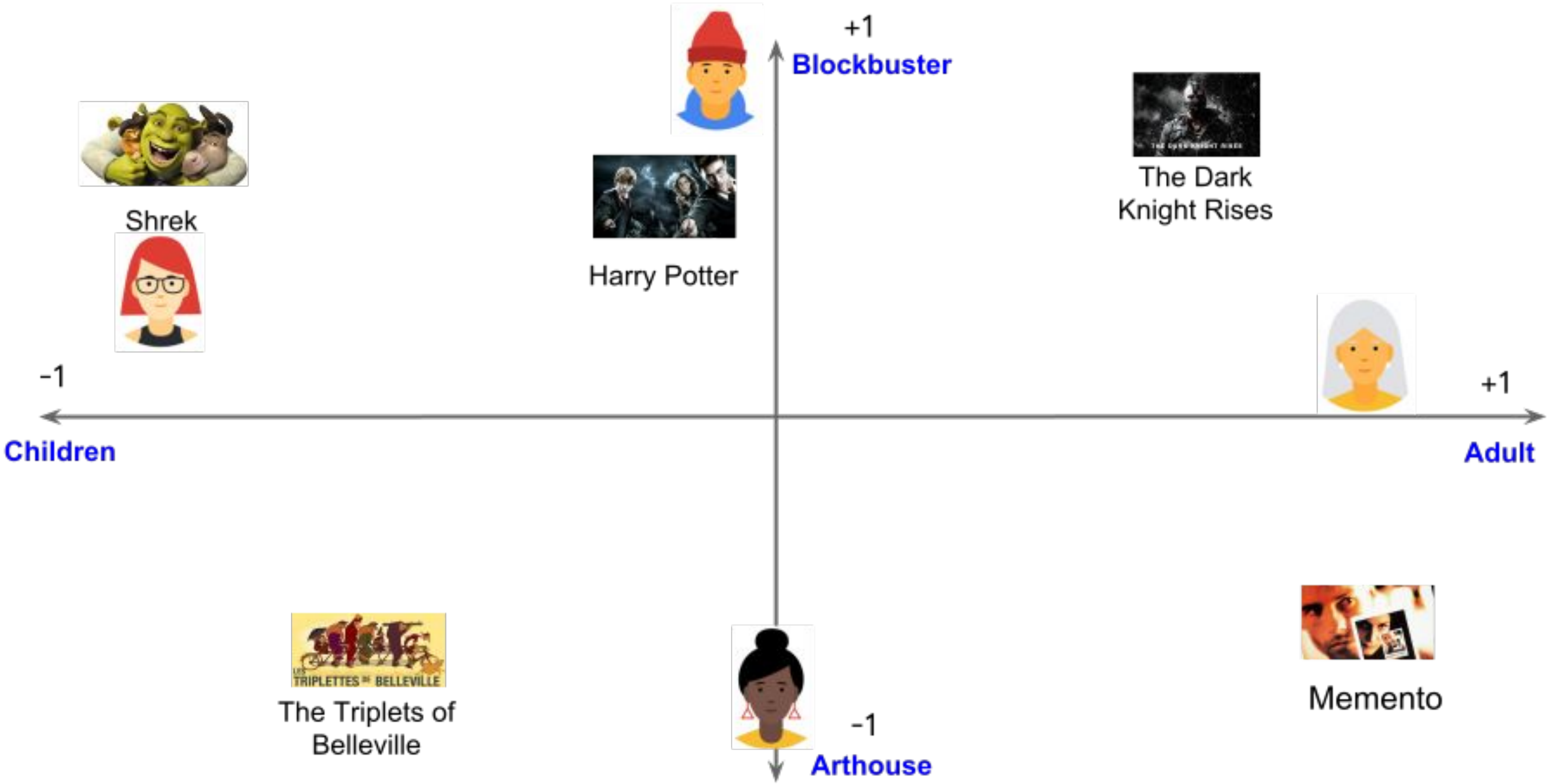
Matrix Factorization based method

Consider a movie recommendation system in which the training data consists of a feedback matrix in which:

- Each row represents a user
- Each column represents an item (a movie)

Let there be two features of an item (movie)

- [Adult, Children] => [-1, 1]
- [Blockbuster, Art] => [-1, 1]



(courtesy Google developers)

	Harry Potter	The Triplets of Belleville	Shrek	The Dark Knight Rises	Memento
	✓		✓	✓	
		✓			✓
	✓	✓	✓		
				✓	✓

$\approx$

$$\begin{array}{cc|ccccc} & & .9 & -1 & 1 & 1 & -.9 \\ & & -.2 & -.8 & -1 & .9 & 1 \\ \hline 1 & .1 & .88 & -1.08 & 0.9 & 1.09 & -0.8 \\ -1 & 0 & -0.9 & 1.0 & -1.0 & -1.0 & 0.9 \\ .2 & -1 & 0.38 & 0.6 & 1.2 & -0.7 & -1.18 \\ .1 & 1 & -0.11 & -0.9 & -0.9 & 1.0 & 0.91 \end{array}$$

(courtesy Google developers)

Matrix Factorization based method

- Two embeddings, mainly User embedding and Movie embedding are learned such that the product UV^T is a good approximation of the feedback matrix A.
- Matrix factorization gives more compact representation
- Finds hidden structure in the data

Drawbacks

- Cannot handle fresh items
- Hard to include side features for query/item - The side features might include country.
- Folding Problem - Unrelated items might be close to relevant items in the embedding space

These lead us to the development of DNN based models.

Deep Neural Network Based

- Easily incorporate query features and item features
- Automatically learn lower dimensional feature mappings
- Better performance that matrix factorization method
- Does not depend on hand-engineered features
- However, computationally expensive, hence embedding layer is used

Deep Neural Network Based

- Utilizing for probabilistic vector generation
- The input to the hidden layers is not the vector input but the embedding layer
- Side features can be incorporated along with embedding input

Deep Neural Network Based - Validation and Testing

- If say, the user watched 20 movies - Train on 14 and holdout 6
- The embedding layer captures low dimensional structure during NN training
- Based on the said 16 movies, the model is evaluated on the ability to predict the holdout set
- Since the comparison is between two probability distributions, Cross entropy loss can be used

Implementation

- Implemented two Content Based Filtering recommendation model
  - Movie Overviews and Taglines
  - Movie Cast, Crew, Keywords and Genre

```
get_recommendations('The Dark Knight').head(10)
```

```
7931                      The Dark Knight Rises
132                              Batman Forever
1113                             Batman Returns
8227    Batman: The Dark Knight Returns, Part 2
7565                 Batman: Under the Red Hood
524                                      Batman
7901                           Batman: Year One
2579              Batman: Mask of the Phantasm
2696                                        JFK
8165    Batman: The Dark Knight Returns, Part 1
Name: title, dtype: object
```

Our Movie Overviews and Taglines Recommender System can distinguish “The Dark Knight” as a Batman film and along these lines suggest other Batman films as its top proposals. This isn't very useful to a great many people as it doesn't take into consideration significant highlights, for example, cast, group, director and classification, which decide the rating and the ubiquity of a film. Somebody who enjoyed The Dark Knight presumably loves it more in view of Nolan and would necessarily enjoy Batman Forever and other films in the Batman Franchise.

```
get_recommendations('The Dark Knight').head(10)
```

```
8031         The Dark Knight Rises
6218                 Batman Begins
6623                  The Prestige
7648                     Inception
2085                     Following
4145                      Insomnia
3381                       Memento
8613                  Interstellar
7659    Batman: Under the Red Hood
1134                Batman Returns
Name: title, dtype: object
```

By using Data like Movie Cast, Crew, Keywords and Genre, We get much more satisfying results this time. The recommendations seem to have recognized other Christopher Nolan movies (due to the high weightage given to the director) and put them as top recommendations. There is a correlation with the likes of The Dark Knight as well as some of the other ones in the list including Batman Begins, The Prestige and The Dark Knight Rises.

Implementation

- Implemented a Deep Learning Model to decompose our higher order feature matrix to two lower order feature matrix.
- Here you can see the top movies that user a has already rated, including the predictions column showing the values the user would have rated. Some of the prediction values seem off (those with value 3.7, 3.8, 3.9 etc.).

	user_id	movie_id	rating	prediction	title	genres
0	2	1357	5	3.487569	Shine (1996)	Drama|Romance
1	2	2236	5	3.621458	Simon Birch (1998)	Drama
2	2	3147	5	4.072145	Green Mile, The (1999)	Drama|Thriller
3	2	1293	5	3.557253	Gandhi (1982)	Drama
4	2	110	5	4.587465	Braveheart (1995)	Action|Drama|War
5	2	3471	5	3.584757	Close Encounters of the Third Kind (1977)	Drama|Sci-Fi
6	2	1945	5	4.091817	On the Waterfront (1954)	Crime|Drama
7	2	1225	5	3.951509	Amadeus (1984)	Drama
8	2	515	5	4.246513	Remains of the Day, The (1993)	Drama
9	2	480	5	3.751309	Jurassic Park (1993)	Action|Adventure|Sci-Fi
10	2	1370	5	3.546409	Die Hard 2 (1990)	Action|Thriller
11	2	1193	5	4.630493	One Flew Over the Cuckoo's Nest (1975)	Drama
12	2	590	5	4.077894	Dances with Wolves (1990)	Adventure|Drama|Western
13	2	1196	5	3.896350	Star Wars: Episode V - The Empire Strikes Back...	Action|Adventure|Drama|Sci-Fi|War
14	2	593	5	4.320173	Silence of the Lambs, The (1991)	Drama|Thriller
15	2	1954	5	4.272499	Rocky (1976)	Action|Drama
16	2	1124	5	3.753161	On Golden Pond (1981)	Drama
17	2	1957	5	3.696649	Chariots of Fire (1981)	Drama
18	2	380	5	3.633754	True Lies (1994)	Action|Adventure|Comedy|Romance
19	2	2501	5	4.552500	October Sky (1999)	Drama

Implementation

- Here you can see a recommendation list of unrated 20 movies sorted by prediction value for the same user.

	movie_id	prediction	title	genres
0	2355	5.860406	Bug's Life, A (1998)	Animation|Children's|Comedy
1	858	5.227665	Godfather, The (1972)	Action|Crime|Drama
2	2804	4.723851	Christmas Story, A (1983)	Comedy|Drama
3	527	4.722781	Schindler's List (1993)	Drama|War
4	47	4.696669	Seven (Se7en) (1995)	Crime|Thriller
5	50	4.690553	Usual Suspects, The (1995)	Crime|Thriller
6	3101	4.682765	Fatal Attraction (1987)	Thriller
7	3897	4.678972	Almost Famous (2000)	Comedy|Drama
8	1218	4.678079	Killer, The (Die xue shuang xiong) (1989)	Action|Thriller
9	1704	4.651975	Good Will Hunting (1997)	Drama
10	2905	4.618134	Sanjuro (1962)	Action|Adventure
11	1136	4.557836	Monty Python and the Holy Grail (1974)	Comedy
12	1148	4.547110	Wrong Trousers, The (1993)	Animation|Comedy
13	926	4.537925	All About Eve (1950)	Drama
14	1254	4.498518	Treasure of the Sierra Madre, The (1948)	Adventure
15	913	4.490264	Maltese Falcon, The (1941)	Film-Noir|Mystery
16	1203	4.448615	12 Angry Men (1957)	Drama
17	1304	4.436367	Butch Cassidy and the Sundance Kid (1969)	Action|Comedy|Western
18	1250	4.427529	Bridge on the River Kwai, The (1957)	Drama|War
19	908	4.425975	North by Northwest (1959)	Drama|Thriller

Thank You