Movie Recommendation System

--- Based on Encoder and K-Means

Zhaoxi Chen | N18963553 | zc1134
Shangwen Yan | N17091204 | sy2160

Abstract

recommender systems have become increasingly popular in recent years, and are utilized in a variety of areas including movies, music, news, books, research articles, search queries, social tags, and products in general. Here we implement a movie recommendation system based on Auto-Encoder and K-Meanse. The dataset is the popular one : movie-lens.

Introduction

With the increasing development of technology and Internet, more and more customers choose to spend their spare time online instead of off-line. As a result, it's becoming of great importance to make a good use of the data produced by the consumers to offer a better consumption environment which is of great benefit both for merchants and customers.
As one of the results, recommender systems have become increasingly popular in recent years, and are utilized in a variety of areas including movies, music, news, books, research articles, search queries, social tags, and products in general. Here we implement a movie recommendation system using movie-lens dataset.

Related works

movie recommendation

Traditional way is to take the user, predict the ratings for all the movies from movies' catalog, sort the movies in descending order by the predicted ratings, then take the top 10 movies and recommend them. Some CNN methods also used in movie recommendation system. But all of this algorithm only focus on genres of movies but not user's characteristics.

Encoder+K-means

This combination actually used more often in image classification. It's main idea is first move redundancy in images by an encoder. And then clustering images with encoded image can speed up the whole algorithm.

Dataset

Original dataset

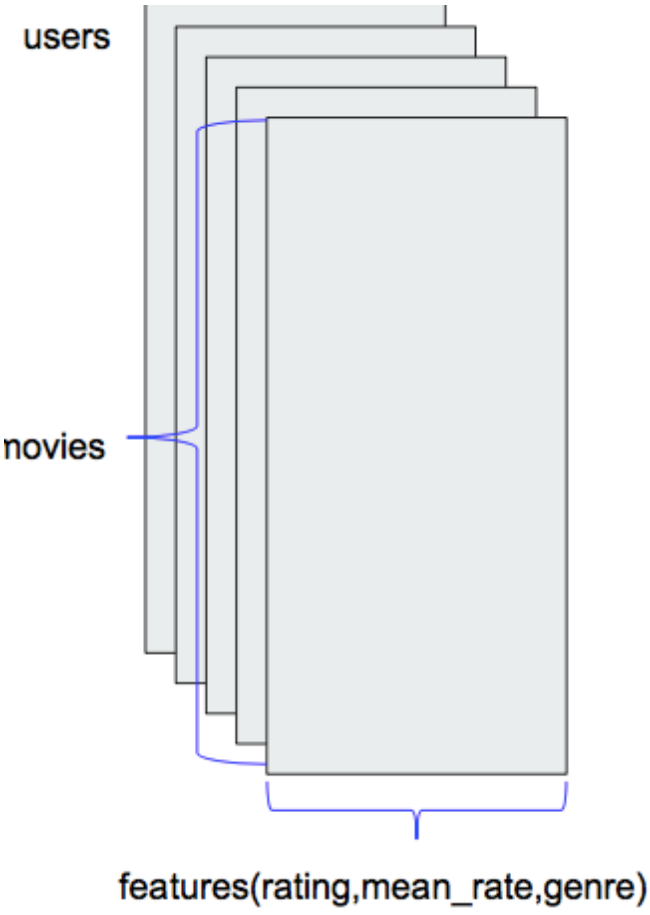
Here we use the open dataset MovieLens. The relevant information is movies'ratings,genres,tags given by users
Number of users:7120， Number of movies: 27278， Number of ratings: 1048575
Files we use:

movie.csv		
movieId	title	genres

rating.csv			
userId	movieId	rating	timestamp

Training data

Since it's a clustering problem, it's more important to train the clusters. We only use 7100 users for traing, and make the data in to a torch of size (7100,27278,3) which means (number of users, number of movies, (rating, mean_rating, genres)). Since genres are given as text, we use one-hot-coding to turn it into binary int, and then decimal int.

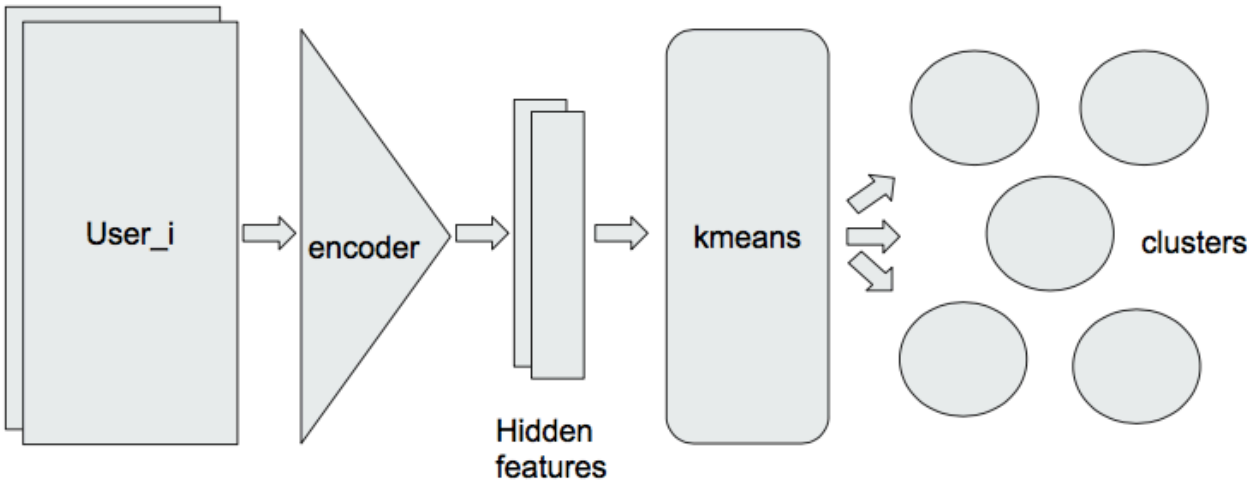


Testing data

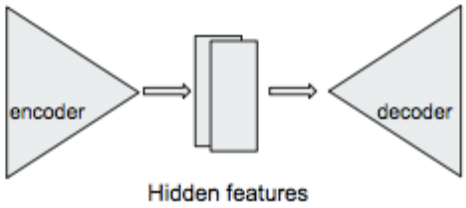
We only use 20 users for testing case. And for those testing users, we split the most recently 20 ratings used to compared with the recommended movies, and the past records to go through the encoder and k-means to do the recommendation

Architecture

The whole architecture is: first we use encoder to reduce the number of features which makes it simpler for the kmeans part, then use k-means to group the users into 9 clusters. After calculating the scores movies in each clusters, we recommend 30 movies which have not been seen by the users to them.



Encoder part



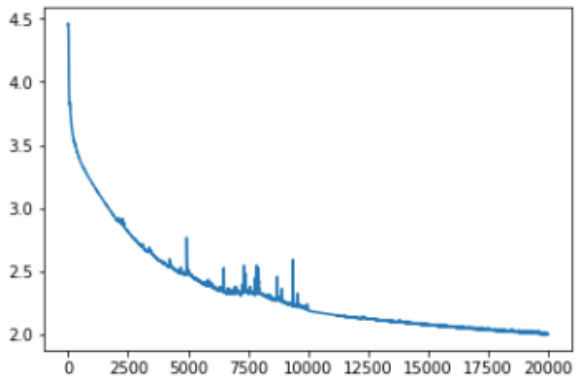
First, we train an autoencoder to make the decoded and input matrix as samilar as possible.

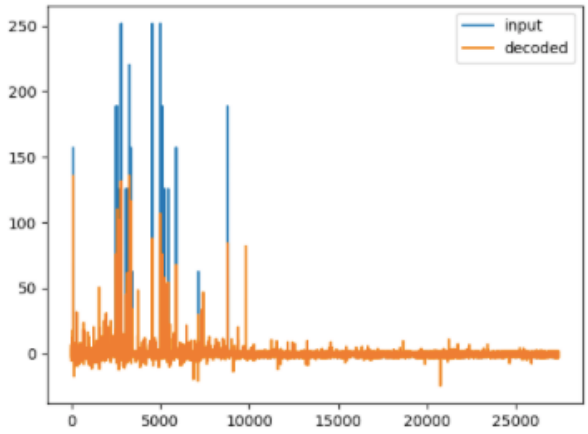
To make sure AE's weights are orthogonal , we add penalty to **MSE loss:**

$L(W) = MSE + \lambda(W^T W - I)$

epoch vs loss:

epoch vs loss	input vs decoded

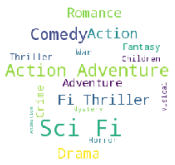
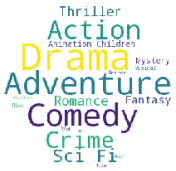
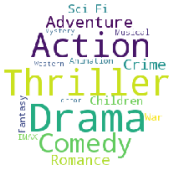
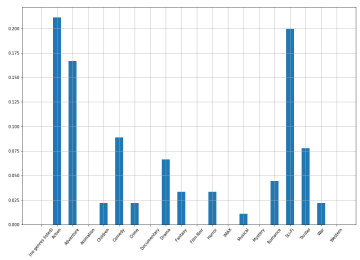
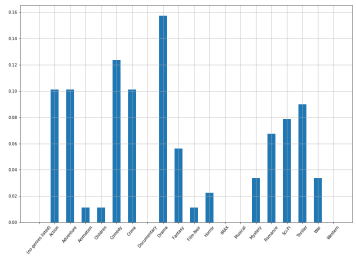
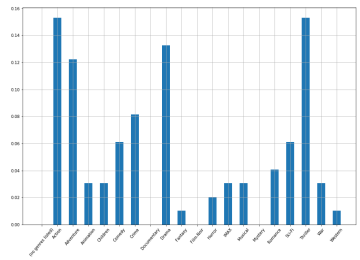
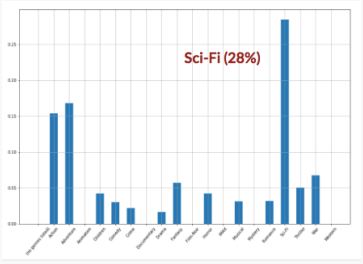
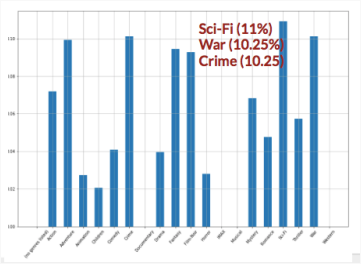
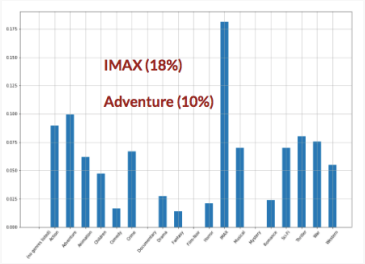


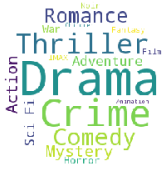
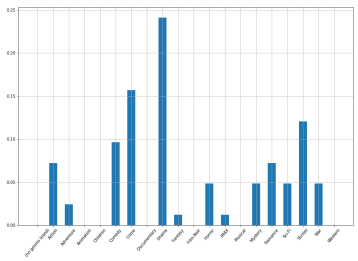
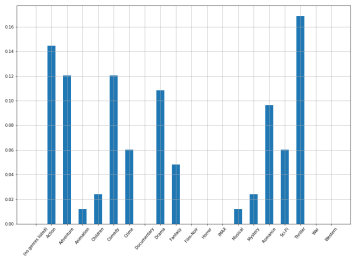
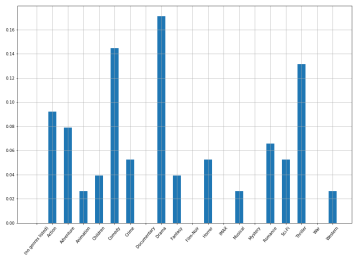
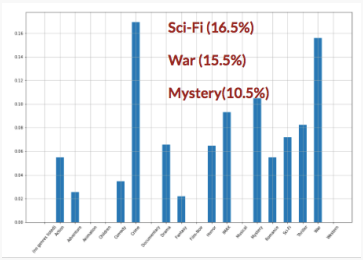
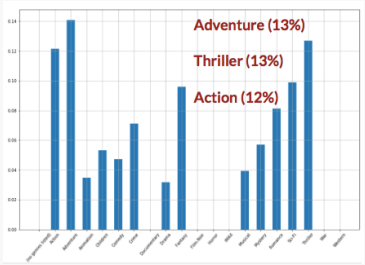
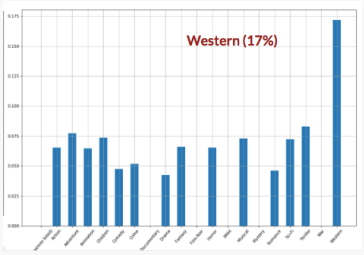
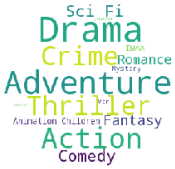
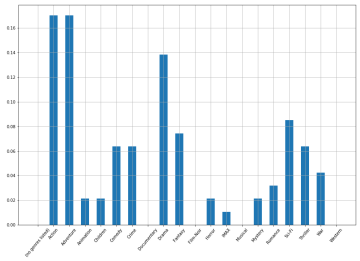
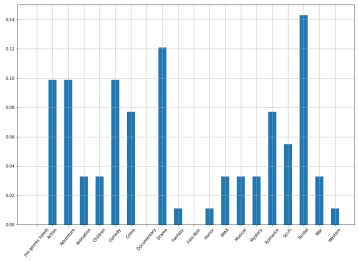
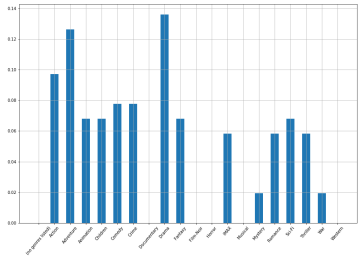
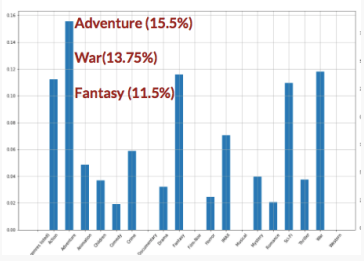
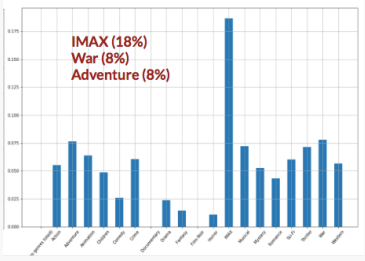
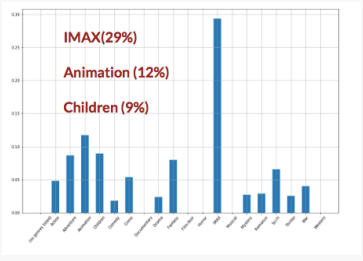


The encoded tensor represents input tensor well, but with fewer features, which makes it easier for clustering in k-means part.

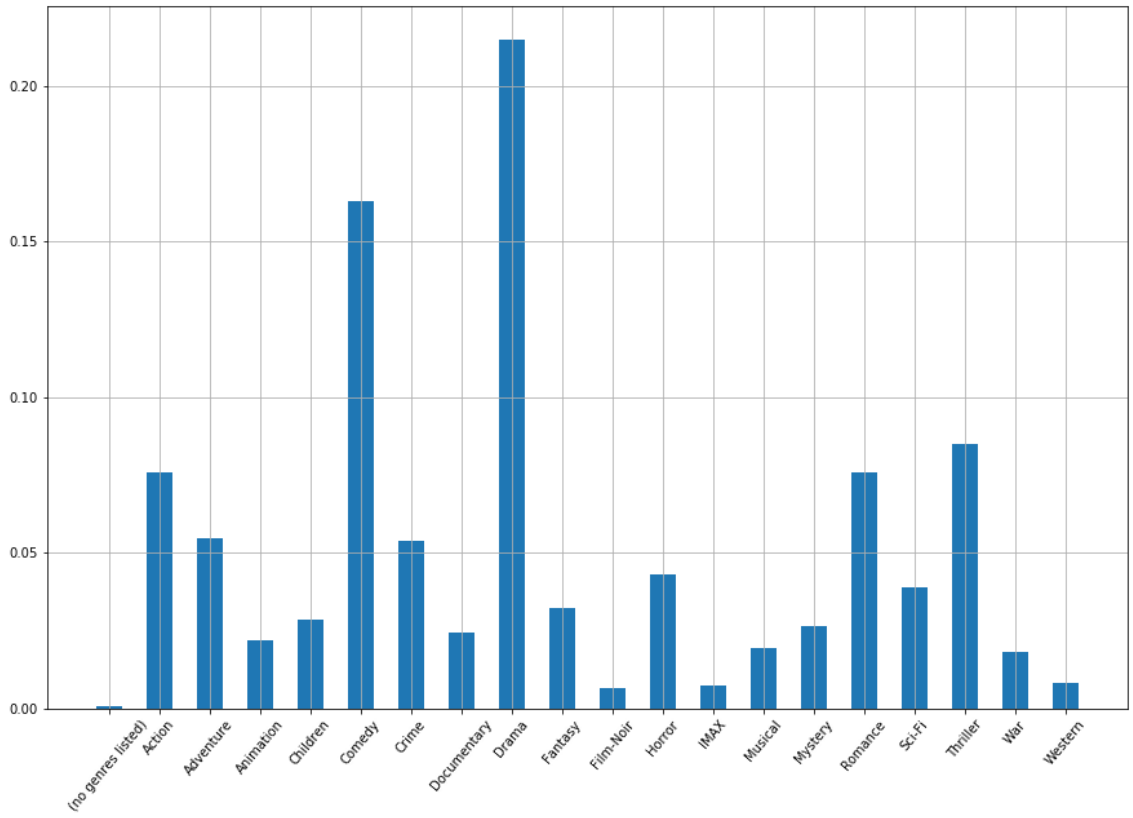
k-means part:

Then use encoded tensor to do a k-means clustering which generates 9 clusters of user-groups.
For each cluster(user_group), we sum up users' ratings for each movie and sort the movies by rating desc.
Then use the top 50 movies' genres to generate a word cloud to see which kind of movie is popular in that user group.

word cloud			
bar plot			
after scaling			

word cloud			
bar plot			
after scaling			
word cloud			
bar plot			
after scaling			

Since ‘Drama’ and ‘Comedy show up frequently, we do a genre analysis among all movies(no clusters).



From the plot we could find some genres, like Drama and Comedy, are ‘common types’ which are specified to most movies. And thus, most clusters contains this two genres is reasonable.
So we scale the frequency of genres based on their probabilities among all movies. As we can see, the difference after scaling is kind of obvious.

Recommen Movies

For every users go through the encoder and kmeans to put him into a cluster. Then recommend those movies with highest scores rated by users in the same cluster to him.

```
================================================================
For user  667
========Recommend Movies=======
Movie name:  Godfather, The (1972)  ||genres:  Crime|Drama
Movie name:  Shawshank Redemption, The (1994)  ||genres:  Crime|Drama
Movie name:  Godfather: Part II, The (1974)  ||genres:  Crime|Drama
Movie name:  American Beauty (1999)  ||genres:  Drama|Romance
Movie name:  Usual Suspects, The (1995)  ||genres:  Crime|Mystery|Thriller
Movie name:  Fargo (1996)  ||genres:  Comedy|Crime|Drama|Thriller
Movie name:  Silence of the Lambs, The (1991)  ||genres:  Crime|Horror|Thriller
Movie name:  Pulp Fiction (1994)  ||genres:  Comedy|Crime|Drama|Thriller
Movie name:  One Flew Over the Cuckoo's Nest (1975)  ||genres:  Drama
Movie name:  Schindler's List (1993)  ||genres:  Drama|War
Movie name:  Fight Club (1999)  ||genres:  Action|Crime|Drama|Thriller
Movie name:  Goodfellas (1990)  ||genres:  Crime|Drama
Movie name:  Shakespeare in Love (1998)  ||genres:  Comedy|Drama|Romance
Movie name:  Taxi Driver (1976)  ||genres:  Crime|Drama|Thriller
Movie name:  Sixth Sense, The (1999)  ||genres:  Drama|Horror|Mystery
Movie name:  Reservoir Dogs (1992)  ||genres:  Crime|Mystery|Thriller
Movie name:  Psycho (1960)  ||genres:  Crime|Horror
Movie name:  Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1964)  ||genres:  Comedy|War
Movie name:  Star Wars: Episode IV - A New Hope (1977)  ||genres:  Action|Adventure|Sci-Fi
Movie name:  Forrest Gump (1994)  ||genres:  Comedy|Drama|Romance|War
Movie name:  Casablanca (1942)  ||genres:  Drama|Romance
Movie name:  Being John Malkovich (1999)  ||genres:  Comedy|Drama|Fantasy
Movie name:  Ghostbusters (a.k.a. Ghost Busters) (1984)  ||genres:  Action|Comedy|Sci-Fi
Movie name:  Life Is Beautiful (La Vita è bella) (1997)  ||genres:  Comedy|Drama|Romance|War
Movie name:  Rear Window (1954)  ||genres:  Mystery|Thriller
Movie name:  Matrix, The (1999)  ||genres:  Action|Sci-Fi|Thriller
Movie name:  Dark Knight, The (2008)  ||genres:  Action|Crime|Drama|IMAX
Movie name:  Good Will Hunting (1997)  ||genres:  Drama|Romance
Movie name:  Star Wars: Episode V - The Empire Strikes Back (1980)  ||genres:  Action|Adventure|Sci-Fi
Movie name:  Blair Witch Project, The (1999)  ||genres:  Drama|Horror|Thriller
========Actually Seen Movies=======
Movie name:  Delta of Venus (1995)  || genres:  Drama
Movie name:  Black Sheep (1996)  || genres:  Comedy
Movie name:  Faces (1968)  || genres:  Drama
Movie name:  Mother (1996)  || genres:  Comedy
Movie name:  2001: A Space Odyssey (1968)  || genres:  Adventure|Drama|Sci-Fi
Movie name:  Time to Kill, A (1996)  || genres:  Drama|Thriller
Movie name:  Cemetery Man (Dellamorte Dellamore) (1994)  || genres:  Horror
Movie name:  Carried Away (1996)  || genres:  Drama|Romance
Movie name:  Liar Liar (1997)  || genres:  Comedy
Movie name:  Mixed Nuts (1994)  || genres:  Comedy
Movie name:  Paris, France (1993)  || genres:  Comedy
Movie name:  Forrest Gump (1994)  || genres:  Comedy|Drama|Romance|War
Movie name:  Mr. Wonderful (1993)  || genres:  Comedy|Romance
Movie name:  House of the Spirits, The (1993)  || genres:  Drama|Romance
Movie name:  Romper Stomper (1992)  || genres:  Action|Drama
Same movies among 30 recommended: 1
================================================================
For user  668
```

Test Case

Recommend 30 movies to a user based on his past watching history. Compare the results with those movies he most recently watched.
here shoes one user's result:

```
==========================================================
For user  661
========Recommend Movies=======
Movie name:  Pulp Fiction (1994)  ||genres:  Comedy|Crime|Drama|Thriller
Movie name:  Shawshank Redemption, The (1994)  ||genres:  Crime|Drama
Movie name:  Silence of the Lambs, The (1991)  ||genres:  Crime|Horror|Thriller
Movie name:  Forrest Gump (1994)  ||genres:  Comedy|Drama|Romance|War
Movie name:  Jurassic Park (1993)  ||genres:  Action|Adventure|Sci-Fi|Thriller
Movie name:  Fugitive, The (1993)  ||genres:  Thriller
Movie name:  Dances with Wolves (1990)  ||genres:  Adventure|Drama|Western
Movie name:  Apollo 13 (1995)  ||genres:  Adventure|Drama|IMAX
Movie name:  True Lies (1994)  ||genres:  Action|Adventure|Comedy|Romance|Thriller
Movie name:  Batman (1989)  ||genres:  Action|Crime|Thriller
Movie name:  Aladdin (1992)  ||genres:  Adventure|Animation|Children|Comedy|Musical
Movie name:  Terminator 2: Judgment Day (1991)  ||genres:  Action|Sci-Fi
Movie name:  Schindler's List (1993)  ||genres:  Drama|War
Movie name:  Braveheart (1995)  ||genres:  Action|Drama|War
Movie name:  Star Wars: Episode IV - A New Hope (1977)  ||genres:  Action|Adventure|Sci-Fi
Movie name:  Stargate (1994)  ||genres:  Action|Adventure|Sci-Fi
Movie name:  Toy Story (1995)  ||genres:  Adventure|Animation|Children|Comedy|Fantasy
Movie name:  Die Hard: With a Vengeance (1995)  ||genres:  Action|Crime|Thriller
Movie name:  Clear and Present Danger (1994)  ||genres:  Action|Crime|Drama|Thriller
Movie name:  Lion King, The (1994)  ||genres:  Adventure|Animation|Children|Drama|Musical|
Movie name:  Beauty and the Beast (1991)  ||genres:  Animation|Children|Fantasy|Musical|
Romance|IMAX
Movie name:  Ace Ventura: Pet Detective (1994)  ||genres:  Comedy
Movie name:  Star Trek: Generations (1994)  ||genres:  Adventure|Drama|Sci-Fi
Movie name:  Seven (a.k.a. Se7en) (1995)  ||genres:  Mystery|Thriller
Movie name:  Usual Suspects, The (1995)  ||genres:  Crime|Mystery|Thriller
Movie name:  Twelve Monkeys (a.k.a. 12 Monkeys) (1995)  ||genres:  Mystery|Sci-Fi|Thriller
Movie name:  Cliffhanger (1993)  ||genres:  Action|Adventure|Thriller
Movie name:  Independence Day (a.k.a. ID4) (1996)  ||genres:  Action|Adventure|Sci-Fi|
Thriller
Movie name:  Speed (1994)  ||genres:  Action|Romance|Thriller
Movie name:  Batman Forever (1995)  ||genres:  Action|Adventure|Comedy|Crime
========Actually Seen Movies=======
Movie name:  Foxfire (1996)  || genres:  Drama
Movie name:  Crow, The (1994)  || genres:  Action|Crime|Fantasy|Thriller
Movie name:  Beauty and the Beast (1991)  || genres:  Animation|Children|Fantasy|Musical|
Romance|IMAX
Movie name:  Punch-Drunk Love (2002)  || genres:  Comedy|Drama|Romance
Movie name:  Infernal Affairs 2 (Mou gaan dou II) (2003)  || genres:  Action|Crime|Drama|
Thriller
Movie name:  Southpaw (2015)  || genres:  Action|Drama
Movie name:  Speed 2: Cruise Control (1997)  || genres:  Action|Romance|Thriller
Movie name:  St. Elmo's Fire (1985)  || genres:  Drama|Romance
Movie name:  Spawn (1997)  || genres:  Action|Adventure|Sci-Fi|Thriller
Movie name:  Sahara (2005)  || genres:  Action|Adventure|Comedy
Movie name:  Inkheart (2008)  || genres:  Adventure|Fantasy
Movie name:  Conan the Barbarian (1982)  || genres:  Action|Adventure|Fantasy
Movie name:  Unbreakable (2000)  || genres:  Drama|Sci-Fi
Same movies among 30 recommended: 1
==========================================================
```

Recommend:

Beauty and the Beast

Speed

Actually watched:

Beauty and the Beast

Speed 2