Movie Recommendations

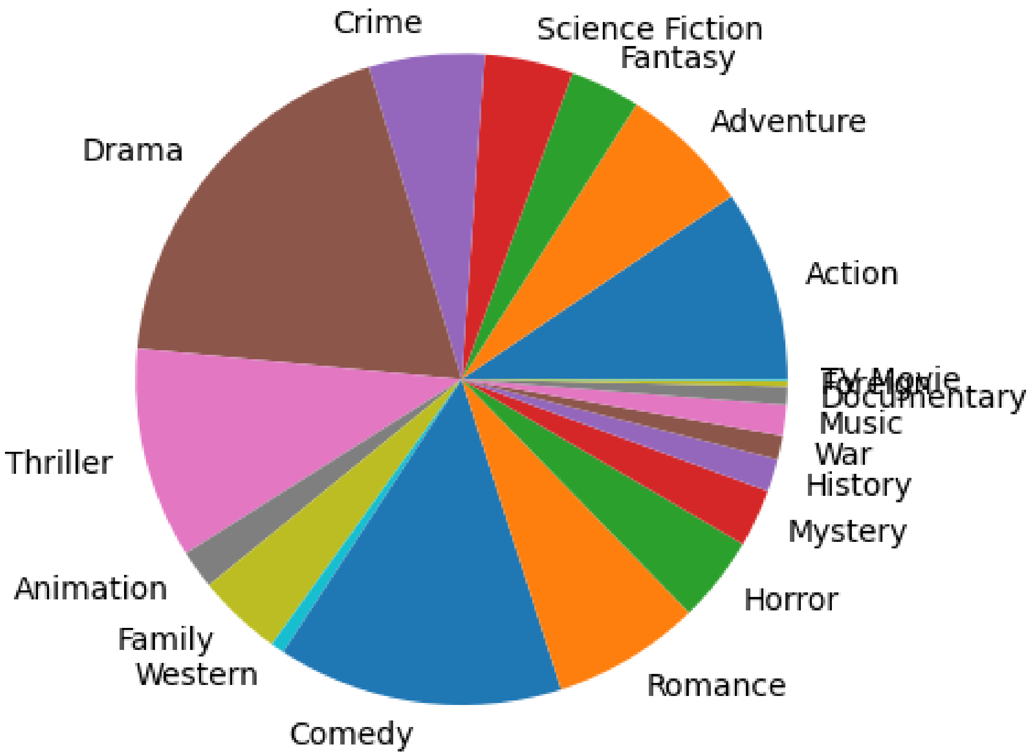
Suggest movie recommendations based on previously liked and disliked films

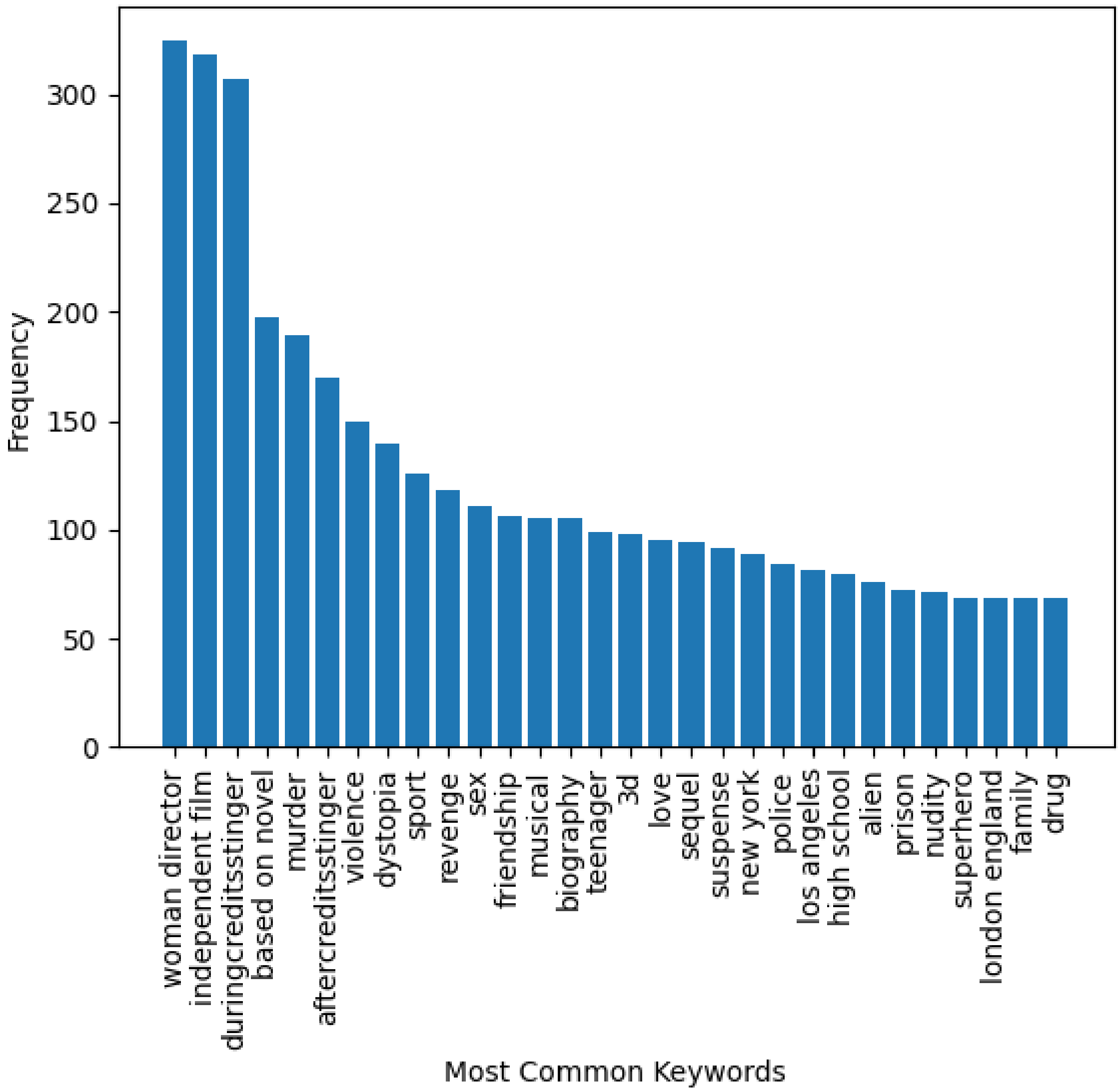
Learning Objectives

- Content-based recommendations
- Bag-of-words
- Collaborative-filtering recommendations

Visualising the Data

A movie database was used, cataloguing 5,000 different films. The database included information about the movies' genres, keywords, title, synopsis, as well as cast and crew.




Frequency
0
50
100
150
200
250
300
woman director
independent film
duringcreditsstinger
based on novel
murder
aftercreditsstinger
violence
dystopia
sport
revenge
sex
friendship
musical
biography
teenager
3d
love
sequel
suspense
new york
police
los angeles
high school
alien
prison
nudity
superhero
london england
family
drug
Most Common Keywords

Bag-of-Words

Bag-of-words is a text model that ignores word order, focussing on the number of occurrences of each word. The movie synopses were encoded using a bag-of-words approach. First, the synopses were broken down into individual words and the occurrence of each word was counted across the database. The 50 most common words are listed below.

the, a, to, and, of, in, his, is, with, her, he, an, for, on, that, their, as, when, by, who, from, but, they, has, are, at, she, into, after, new, one, young, out, him, up, it, life, be, two, this, find, have, must, all, about, world, man, story, will, only

Most of these most common words give very little insight into the content of the film, so these words were disregarded. The next 10,000 most common words were included in the bag-of-words model.

Movie Encoding

Each movie was encoded by storing its genres, keywords, and synopsis words as separated bags-of-words. For simplicity, each word was either present or absent, rather than counting the number of occurrences.

Movie Similarity

A pair of movies can be assigned a similarity scored based on how many genres, keywords, and synopsis words they have in common. The genre was weighted the most heavily, and the synopsis words were weighted as the inverse of the number of words in the synopsis to prevent skewing the data with very long or very short synopses.

To sanity-check this procedure, the top 10 movies most similar to Batman Begins and Skyfall were generated. They are tabulated below, along with their similarity scores.

Batman Begins		**Skyfall**	
15.1	The Dark Knight	9.0	The Spy Who Loved Me
13.0	The Dark Kight Rises	8.1	Quantum of Solace
9.0	Batman & Robin	8.0	Never Say Never Again
8.1	Batman Returns	8.0	Mission: Impossible
8.1	Batman Forever	8.0	Dr. No
8.0	Teenage Mutant Ninja Turtles	8.0	On Her Majesty’s Secret Service
8.0	Defendor	8.0	Diamonds Are Forever
8.0	Blade	7.1	Spectre
7.1	Brick Mansions	7.1	MI: Ghost Protocol
7.0	Batman v Superman	7.0	Tomorrow Never Dies

Content-Based Recommendation

The movie similarity scores were used to recommend films based on a set of liked and disliked films. For each new movie, their similarity scores with the disliked films were subtracted from their similarity scores with the liked films, giving an overall recommendation score.

Liked: Batman Begins, The Dark Knight, Memento, Inception, Alien
Disliked: The Dark Knight Rises, The Prestige, Cinderella

Recommended Movie	Recommendation Score
Green Lantern	21.0
Pandorum	20.2
Blade: Trinity	20.1
Aliens	20.0
Alien^3	19.1
Starship Troopers	19.1
The Matrix Reloaded	19.1
X2	19.1
Planet of the Apes	19.1
Serenity	19.0

This recommendation algorithm can be refined by splitting the films into more refined categories, such as using a 5-star rating system. Films receiving 1 and 2 stars have a negative weighting (with 1 star having the lowest), whilst films receiving 3, 4, and 5 stars have progressively more positive weightings.

1 Star: The Prestige, The Hobbit: The Battle of the Five Armies
2 Stars: Lost in Translation, The Wizard of Oz
3 Stars: Batman Begins, Dead Poets Society
4 Stars: Persepolis, Up
5 Stars: Memento, The Theory of Everything

Recommended Movie	Recommendation Score
Anomalisa	23.0
Outside Providence	21.5
Inside Out	20.6
Confessions of Dangerous Mind	20.1
Brooklyn's Finest	20.0
Arthur Christmas	20.0
When Harry Met Sally...	19.0
Ernest et Celestine	19.0
Goodbye, Lenin!	18.8
Home Fries	18.4

Collaborative-Filtering Recommendation

Collaborative-filtering recommendations do not consider the properties of the movies or the preferences of the individual user. Instead, they leverage the feedback and history of all users, inferring correlations in preferences between users with similar histories. This is the “customers who bought this item also bought…” approach to recommendation.

The advantage of collaborative-filtering is its ability to harness a large dataset of preferences, allowing it to produce more accurate results than content-based methods. However, it requires many pre-existing user histories in order to fit a reliable model.

Suppose the movies represent a network, where each node is a different film. An edge is added between two movies if a user likes both movies. The more users who like both films, the more edges that will connect these nodes. Highly connected regions of the network suggest similar films, allowing recommendations to be made.

Conclusion

Content-based recommendation algorithms can suggest new movies to watch based on prior liked / disliked movies. A new movie is suggested if it has a high similarity to the liked films and a low similarity to the disliked films. Bag-of-words is an effective tool to estimate the similarity of two films by comparing overlapping words in the movies' genres, keywords, and synopses.

Collaborative-filtering recommendation algorithms infer user preferences from data from similar users, producing more reliable recommendations but requiring a large pre-existing database of user preferences.