Building Your First Recommendation Algorithm

Devfest 2024

Goals & Guiding Questions

Let’s design an algorithm that can give good movie recommendations!

Let’s apply what we learned today to make cool things for the hackathon!

Goals & Guiding Questions

1. What makes a program intelligent?

2. How can we represent movie data?

3. How can we define movie similarity?

4. How can we define “good” recommendations?

5. How can we implement this in code?

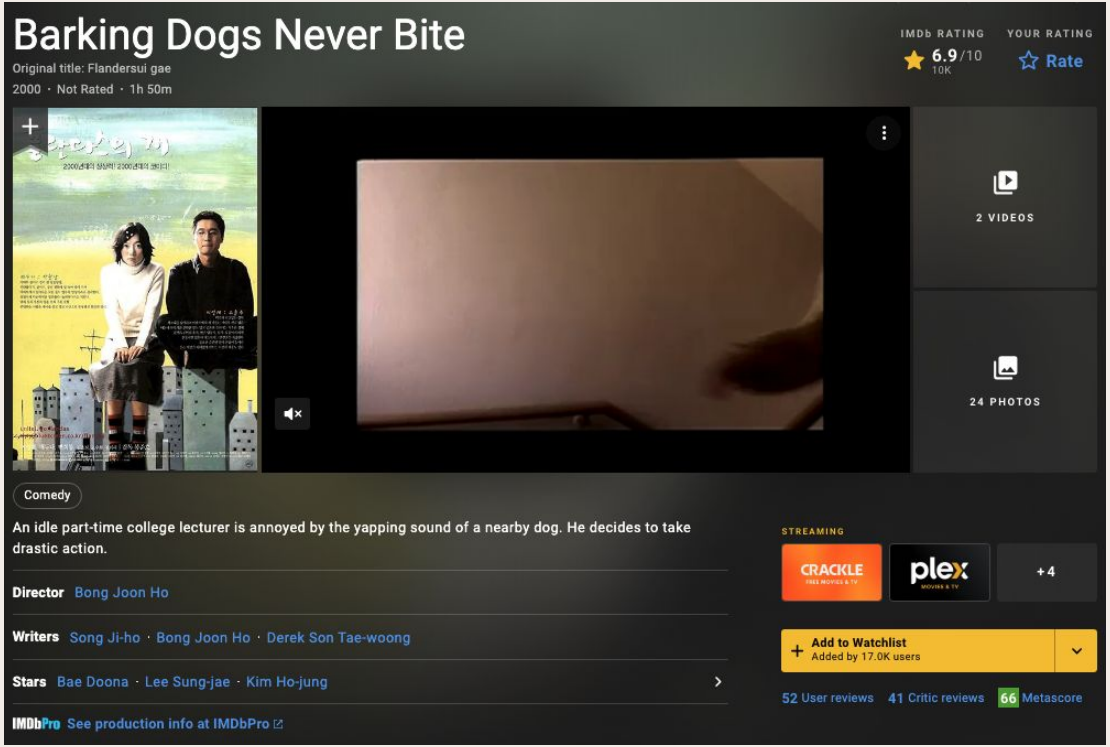
6. How can I use this for the hackathon?

1. What makes a program “intelligent”?

Adaptability

Learning Context

2. How can we represent movie data?



2. How can we represent movie data?

MOVIE INFO

Frustrated with loud barking, an academic (Lee Sung-jae) wages war against dogs in his apartment building.

Content collapsed.

Genre: Comedy

Original Language: Korean

Director: Bong Joon Ho

Writer: Bong Joon Ho, Tae-woong Derek Son, Song Ji-ho

Release Date (Streaming): Nov 1, 2011

Runtime: 1h 48m

Production Co: CJ Entertainment, Uno Films, Cinema Service

2. How can we represent movie data?

Is the data a good representation of the actual object?

Is the data easily storable and perform functions on?

2. How can we represent movie data?

Use the entire script

Take the plot summary

Set of keywords

Overlapping actors / directors

Genre metrics

https://en.wikipedia.org/wiki/Bag-of-words_model

2. How can we represent movie data?

Comedy: 9.53
Romance: 1.01
Action: 1.22
Thriller: 2.33

...

Mystery: 0.87
Suspense: 8.09

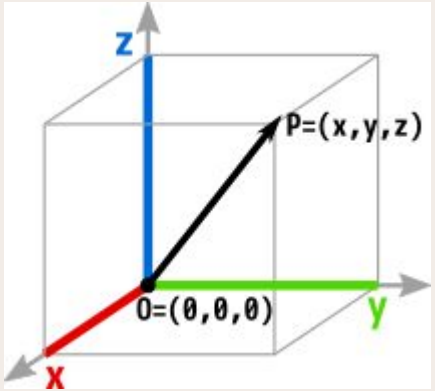
[9.53, 1.01, 1.22, 2.33 ... 0.87, 8.09]

3. How can we define movie similarity?

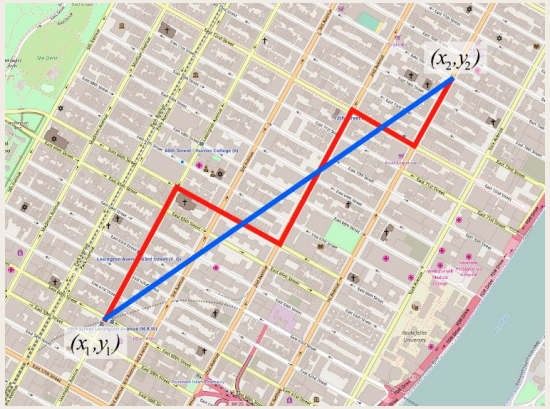
In other words, how should we measure the distance between two points in space?

3. How can we define movie similarity?

Euclidean Distance



Manhattan Distance



4. How can we define good recommendations?

Good recommendation
=
Closest movie?

https://en.wikipedia.org/wiki/Netflix_Prize

5. How can we implement this in code?

Movies I've liked:

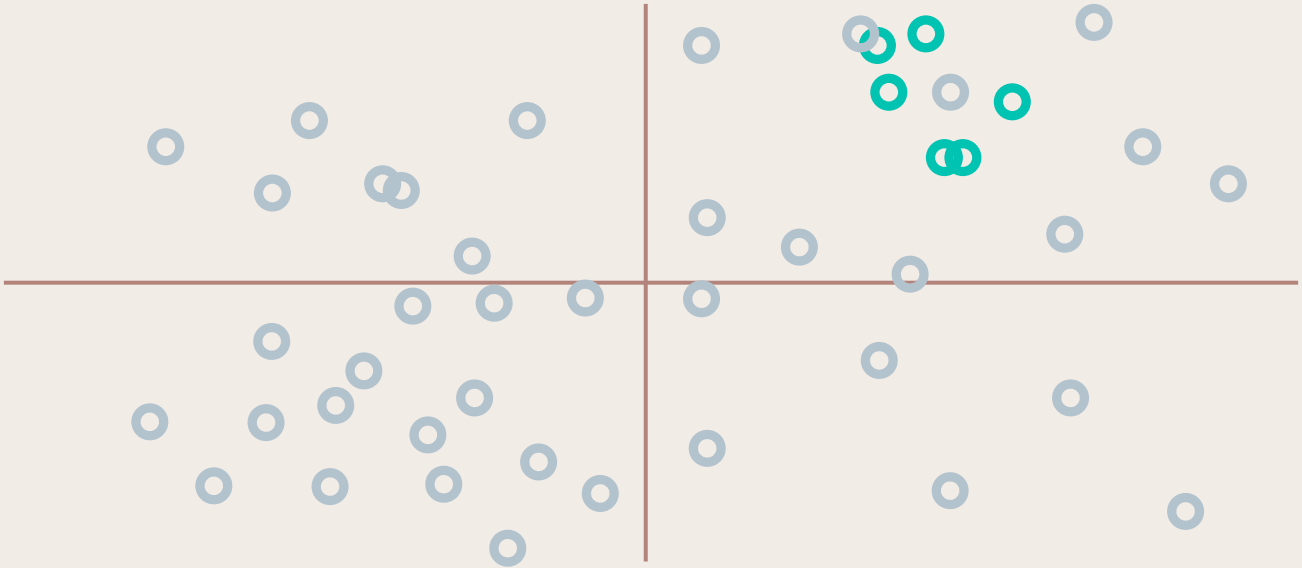
[[0.11, 7.32 ... 3.46],
[2.37, 9.11 ... 3.32],
...
[9.88, 0.12 ... 3.45]]

Our recommendation algorithm

Recommended movie

https://en.wikipedia.org/wiki/Netflix_Prize

5. How can we implement this in code?



https://en.wikipedia.org/wiki/K-means_clustering

5. How can we implement this in code?

```
closest_distance = infinity
recommendation = None

for movie M in all_movies:
    current_distance = 0

    for movie L in liked_movies:
        current_distance += Euclidean(M, L)

    if current_distance < closest_distance:
        recommendation = M

return recommendation
```

https://en.wikipedia.org/wiki/K-means_clustering

5. How can I use this for the hackathon?

1. Make a recommendation algorithm (with improvements!)

2. Dig deeper and explore other applications of AI / ML

3. Come talk to us with your ideas!

4. *** Come hack on Saturday!!

Connect with me!

jyh2134@columbia.edu

https://www.linkedin.com/in/joshuayhahn/