Recommender Systems (7.05 & 7.06)

What are some examples of recommendation engines that you have used before?

Recommender System – Overview

- **Recommender systems** are an important class of machine learning algorithms that offer “relevant” suggestions to users.

- YouTube, Amazon, Netflix, all function on recommendation systems where the system recommends you the next video or product based on your past activity (**Content-based Filtering**) or based on activities and preferences of other users similar to you (**Collaborative Filtering**).

- Likewise, Facebook also uses a recommendation system to suggest Facebook users you may know offline.

Recommender System – Netflix

- In 2000, Netflix introduced personalised movie recommendations and in 2006, launched **Netflix Prize,** a machine learning and data mining competition with a $1 million dollar prize money.

- Back then, Netflix used *Cinematch*, its proprietary recommender system which had a ***root mean squared error*** **(RMSE) of 0.9525** and challenged people to beat this benchmark by 10%. The team who could achieve the target or got close to this target after a year would be awarded the prize money.

Recommender System – Netflix

- The winner of the Progress Prize a year later in 2007 achieved a RMSE of 0.88. Netflix then put those algorithms into production after some adaptations to the source code.

- What is worth noting is that despite some teams achieving a RMSE of 0.8567 in 2009, the company did not put those algorithms into production due to the engineering effort required to gain the marginal increase in accuracy. This serves an important point in real-life recommender systems — that there is always a positive relationship between model improvements and engineering efforts.

Recommender System – Netflix

- A more important reason why Netflix did not incorporate the improved models from the Netflix Prize is because it introduced streaming in 2007.

- With streaming, the amount of data it has surged dramatically.

- It has to change the way its recommender system was generating recommendations and ingesting data.

Recommender System – Netflix

- Fast forward to 2020, Netflix has transformed from a mail service posting DVDs in the US to a global streaming service with 182.8 million subscribers.

- Consequently, its recommender system transformed from a regression problem predicting ratings to a ranking problem, to a page-generation problem, to a problem maximising user experience (defined as maximising number of hours streamed i.e. personalising everything that can be personalised).

Recommender System – Netflix

Source: [Recent Trends in Personalization — A Netflix Perspective]

Recommender System – Netflix

- Netflix has a subscription-based model. Simply put, the more *members* (the term used by Netflix, synonymous to users/subscribers) Netflix has, the higher its revenue.

- Revenue can be seen as a function of three things:
  - Acquisition rate of new users
  - Cancellation rates
  - Rate at which former members rejoin

How important is Netflix's Recommender System?

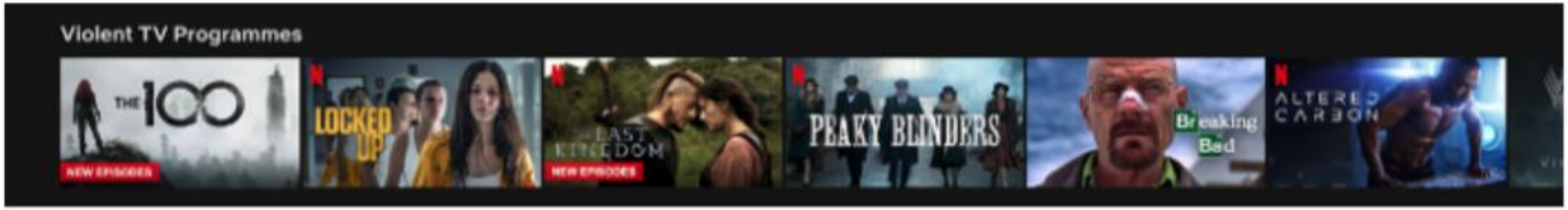
- 80% of stream time is achieved through Netflix's recommender system

- Netflix believes in creating a user experience that will seek to improve retention rate, which in turn translates to savings on customer acquisition (estimated $1B per year as of 2016).

Recommender System – Netflix

What algorithms are used by Netflix?

- **Personalised Video Ranking (PVR)** — This algorithm is a general-purpose one, which usually filters down the catalog by a certain criteria (e.g. Violent TV Programmes, US TV shows, Romance, etc), combined with side features including user features and popularity.

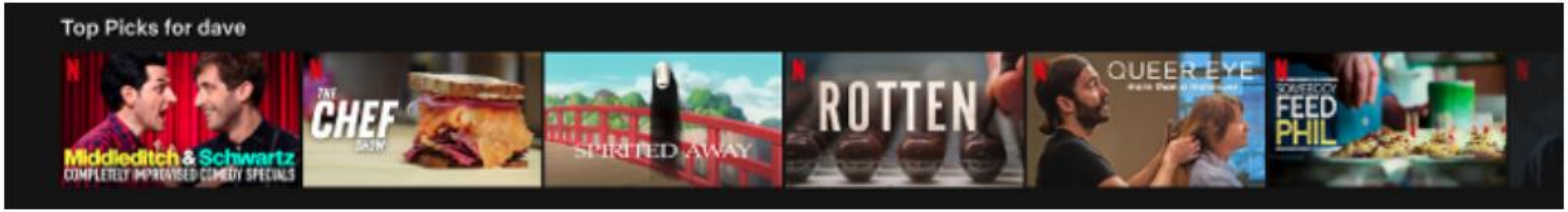


Example of PVR generated items

Recommender System – Netflix

What algorithms are used by Netflix?

- **Top-N Video Ranker** — Similar to PVR except that it only looks at the head of the rankings across the entire catalog.

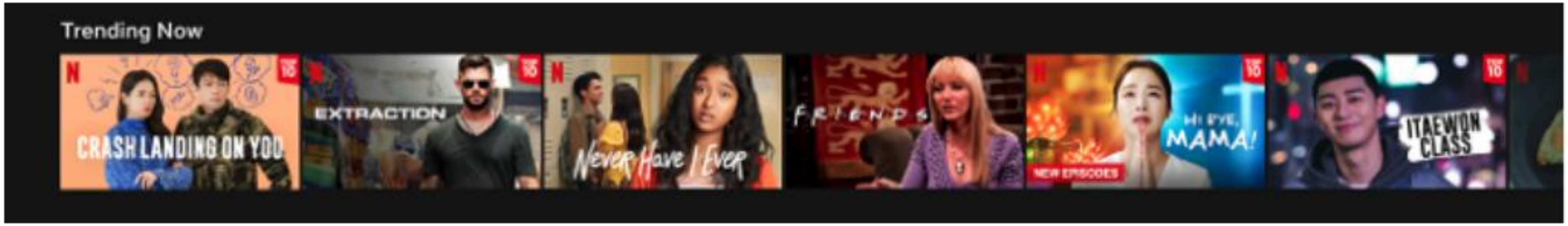


Example of Top-N ranker generated titles

Recommender System – Netflix

- **Trending Now Ranker** — This algorithm captures temporal trends which Netflix deduces to be strong predictors. These short-term trends can range from a few minutes to a few days. These include:

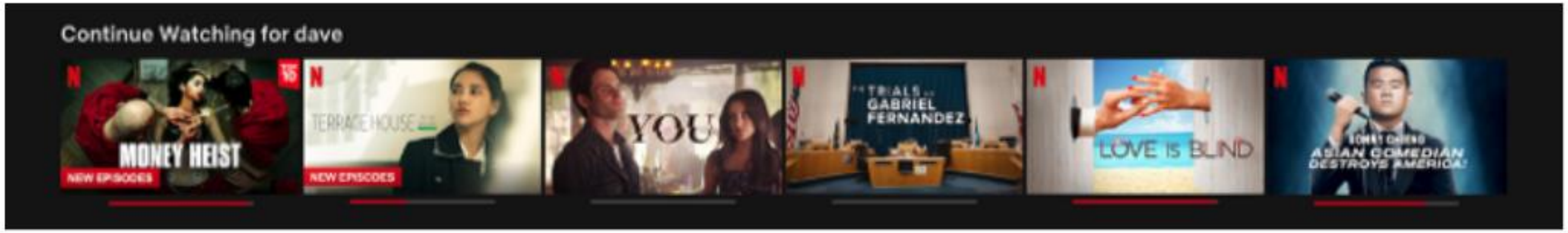
  1. Events that have a seasonal trend and repeat themselves (e.g. Valentines day leads to an uptick in Romance videos being consumed)
  2. One-off, short term events (e.g. Coronavirus or other disasters, leading to short-term interest in documentaries about them)



Example of Trending Now ranker generated titles

Recommender System – Netflix

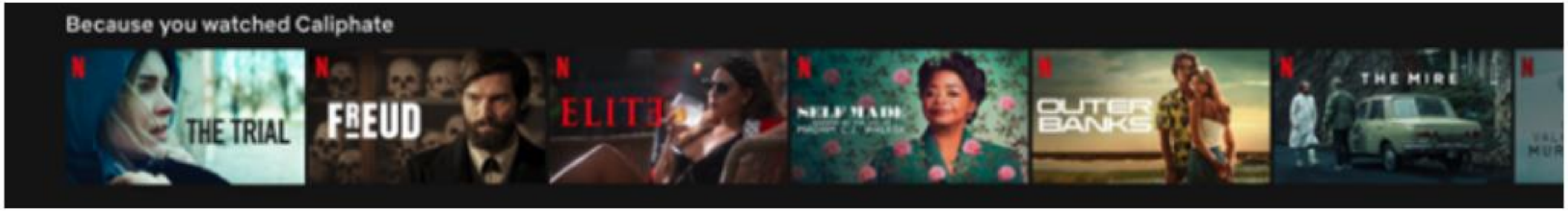
- **Continue Watching Ranker** — This algorithm looks at items that the member has consumed but has not completed, typically:
  - Episodic content (e.g. drama series)
  - Non-episodic content that can be consumed in small bites (e.g. movies that are half-completed, series that are episode independent such as Black Mirror)
- The algorithm calculates the probability of the member continue watching and includes other context-aware signals (e.g. time elapsed since viewing, point of abandonment, device watched on, etc).



Example of Continue Watching ranker generated titles

Recommender System – Netflix

- **Video-Video Similarity Ranker** —a.k.a. Because you watched (BYW)
- This algorithm basically resembles that of a content-based filtering algorithm. Based on an item consumed by the member, the algorithm computes other similar items (using an item-item similarity matrix) and returns the most similar items.



Example of BYW generated titles

Types of Recommenders

- Content-based

- Collaborative
  - User-based
  - Item-based

Content-Based Recommenders

Based on **product features**

Title	Year	Genre	Director	MPAA Rating
Elf	2003	Christmas/Comedy	Jon Favreau	PG
Die Hard	1988	Christmas/Action	John McTiernan	R
How to Train Your Dragon	2010	Animation/Action	Dean DeBlois Chris Sanders	PG

Collaborative Recommenders: User Based

Recommendations from **users** with similar **ratings**

	Movie 1	Movie 2	Movie 3
User 1	5	1	NOT WATCHED
User 2	4	1	4
User 3	1	5	1

Example: Recommend Movie 3 to User 1 since he/she has given the same rating as User 2 for Movie 2

Collaborative Recommenders: User Based

Recommendations from **products** with similar **ratings/purchases**

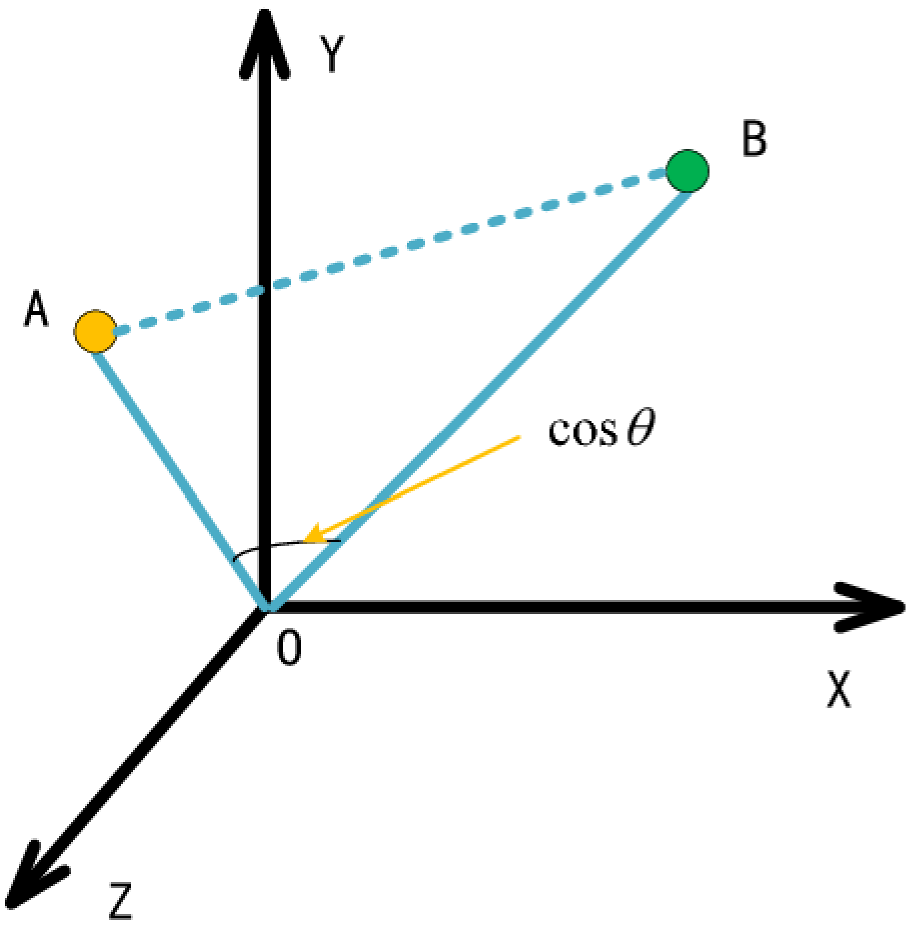
	User 1	User 2	User 3
Movie 1	4	5	2
Movie 2	2	1	5
Movie 3	5	4	2

Example: Recommend Movie 3 to users who have watched Movie 1 but not Movie 3 since Movie 3 has similar set of ratings as Movie 1

Cosine Similarity & Cosine Distance

- Cosine similarity is a metric used to measure how similar two items are.
- Mathematically, it measures the cosine of the angle between two vectors projected in a multi-dimensional space.
- The output value ranges from **0–1** where **0 means no similarity, where as 1 means that both the items are 100% similar.**

Cosine Similarity & Cosine Distance



Similarity = Cos θ

Distance = 1 – Similarity

Higher the similarity, Lower the distance

Exercise – Which user is most similar to Jerome? And what should we recommend next to Jerome?