2021 Fast.ai Community Course

Arun Prakash & Huyen Nguyen

QUEENSLAND HUB

QUEENSLAND AI

Lesson 6 – Recommendation Systems

Notebook: 08_collab.ipynb

About us

- Arun Prakash
- https://twitter.com/arunprakashml
- https://www.linkedin.com/in/arunprakashtce/

Agenda

- Intro to Recommendation Systems
- Code walkthrough
- Datablock, TabularPandas
- Weights and biases (wandb)
- Paperswithcode + using colabcode
- Streamlit
- Tips on projects (Arun won first prize in previous Brisbane fastai course)
- Useful resources on recsys
- MovieLens case study
- Exercise walkthrough and solution

How to find a needle in a haystack?

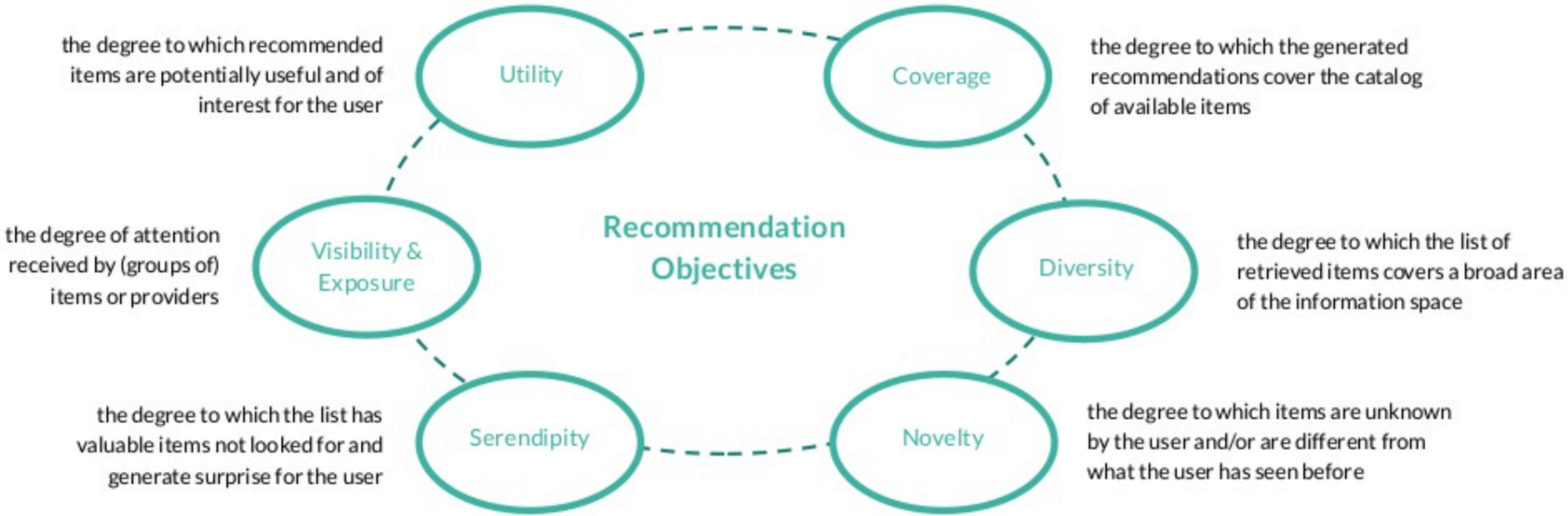
- Problem: We are inundated by a vast amount of data, causing us to be unable to find out our target information even if we have access to it.
- Have you ever tried browsing YouTube on a blank slate?
- Recommendation systems are developed to solve this problem of information overload.
- What ***should*** a recsys do?

Accuracy might be only one of many goals

- The recommendations that are most accurate according to the standard metrics are sometimes not the recommendations that are most useful to users.
- Being accurate is not enough: how accuracy metrics have hurt recommender systems

- A good recommender system has to consider how users interact with the recommendations.
- The wonderful world of recommender systems (Yanir Seroussi, Lesson 5 mentor)

What are we optimizing?



What can go wrong?

What can go wrong?

How Facebook Groups Are Being Exploited To Spread Misinformation, Plan Harassment, And Radicalize People

Mark Zuckerberg wants to get a billion people in "meaningful" Facebook groups. But to get there he'll have to battle the spammers, hackers, and trolls who exploit and hijack groups to make money or sow chaos.

Craig Silverman, BuzzFeed News Media Editor
Jane Lytvynenko, BuzzFeed News Reporter
La… Bu…

Posted on March 19, 2018, at 1:36 p.m. ET

Science · Contents · News · Careers · Journals

Read our COVID-19 research and news.

SHARE

REPORT

The spread of true and false news online

Soroush Vosoughi[1], Deb Roy[1], Sinan Aral[2,*]
+ See all authors and affiliations

Science 09 Mar 2018:
Vol. 359, Issue 6380, pp. 1146-1151
DOI: 10.1126/science.aap9559

Article · Figures & Data · Info & Metrics · eLetters · PDF

Lies spread faster than the truth

There is worldwide concern over false news and the possibility that it can influence political, economic, and social well-being. To understand how false news spreads, Vosoughi *et al.* used a data set of rumor cascades on Twitter from 2006 to 2017. About 126,000 rumors were …d by ~3 million people. False news reached more people than the truth; the top 1% of …ews cascades diffused to between 1000 and 100,000 people, whereas the truth rarely …d to more than 1000 people. Falsehood also diffused faster than the truth. The degree …elty and the emotional reactions of recipients may be responsible for the differences …ed.

What is popular is not necessarily good

The New York Times

Opinion

YouTube, the Great Radicalizer

By Zeynep Tufekci

March 10, 2018

When Recommendations Systems Go Bad

Fake News and Rabbit Holes: radicalization via the recommendation engine

Renee DiResta · Follow
Nov 13, 2016 · 4 min read

Do recommendation systems always lead to a filter bubble?

- Recall the MovieLens data in the notebook
- Do recommender systems expose users to narrower content over time?
- Do users taking the recommendations have a more enjoyable experience than those ignoring them?
- Find out at the end of class

Types of Recommendation Systems

- **Demographic:** recommend based on user attributes such as gender, age, nationality
- **Knowledge-based:** recommend items based on specific domain knowledge about how certain item features meet user preferences
- **Community-based:** recommend items based on the ratings and preferences of users' social network
- What are the pros and cons?

Content-based recsys

- Based on attributes of the item. Recommends items with features/attributes similar to what user already liked in the past.
- Works well when descriptive data on the content is provided beforehand.
- What are the pros and cons?

Content-based recsys

- Pros: Works even when a product has no user reviews
- Cons: Requires descriptive data of all content to recommend
- ‘Features’ are human designed, what is relevant?

Collaborative filtering

- Makes suggestions based on how similar users rated in the past and not based on the product themselves.
- What are the pros and cons?
- How do we measure ‘similarity’?
- Can we represent each user as a vector of ratings and calculate their correlation matrix as a measure of similarity?

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

Why use embeddings?

- Curse of dimensionality: data is sparse. Most user-movie rating combinations do not exist.
  - Most users have only rated a small fraction of movies available.
  - Most movies have only been rated by a small fraction of users.
- Representing users and movies as vectors of mostly 0 is problematic.
- Using embeddings is a dimensionality reduction technique.

MovieLens case study

- Users of MovieLens system who took recommendations from the system narrowed their movie consumption over time. But users who didn't take recommendations from the system narrowed their consumption even more!
- In other words, most people tend to narrow their consumption as:
    - they get to know their tastes better
    - they run out of "blockbuster" items that appeal across genres

Taking recommendations *lessened* the risk of a filter bubble!

[Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity]

User experience of taking vs ignoring recommendations

- “Overall, our users watched less enjoyable movies.”, but…
- On average, the Following Group’s enjoyment remained the same. But they liked the recommended movies better than the non-recommended ones.
- The Ignoring Group’s enjoyment went down over time and watched significantly less enjoyable movies than the Following Group.

Simpson’s paradox

Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity

Recommendations for a good RecSys

- First, they can use collaborative filtering algorithms like those in MovieLens, which slows the narrowing effect over time.
- Second, recommender systems can inform users about the diversity of their consumption. Be it movies or news, a site can display diversity metrics or summary statistics that help users better understand if they have in fact gone too far into a particular interest of theirs.
- Finally, if recommenders aren't enough to reduce the narrowing effect, we should explore further steps to intentionally increase diversification of recommendation lists.

- Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity

DataBlock, TabularPandas

https://ohmeow.com/posts/2020/04/11/finding-datablock-nirvana-part-1.html (From Wayde)

https://www.youtube.com/watch?v=NzWadB_fcTE (From Vishnu, JarvisLabs)

https://docs.fast.ai/collab.html

Coding

https://colab.research.google.com/drive/1GSJngosrY1bMVXsy2T_d7QUqJqnKUvGS?usp=sharing

Colabcode, paperswithcode

https://colab.research.google.com/github/abhishekkrthakur/colabcode/blob/master/colab_starter.ipynb#scrollTo=M6bGiCAGxilQ

https://paperswithcode.com/task/recommendation-systems

https://paperswithcode.com/datasets?q=&v=lst&o=match&task=recommendation-systems

Weights and Biases