Embedded Ethics Module Recommender System Objectives

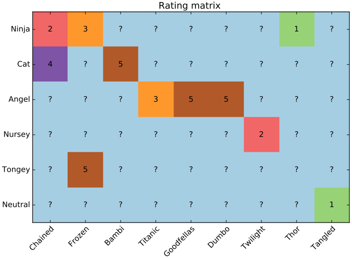
Today

- **Topic:** objective functions for recommender systems
- Two parts
  - **Part 1:** technical challenges in moving beyond regression and classification
  - **Part 2:** ethical challenges, and philosophical tools for reasoning about them

Recap and Motivation

Recap: Netflix Challenge

- We can view collaborative filtering as a matrix completion problem.



- In addition to the learning algorithm, it is important to consider the data and the objective function.

Recommender Systems

- Other kinds of recommendation systems include search engines and social media feeds.
- What are some difficulties you'd run into if you tried to use a Netflix-style algorithm to organize a user's social media feed?

Recommender Systems

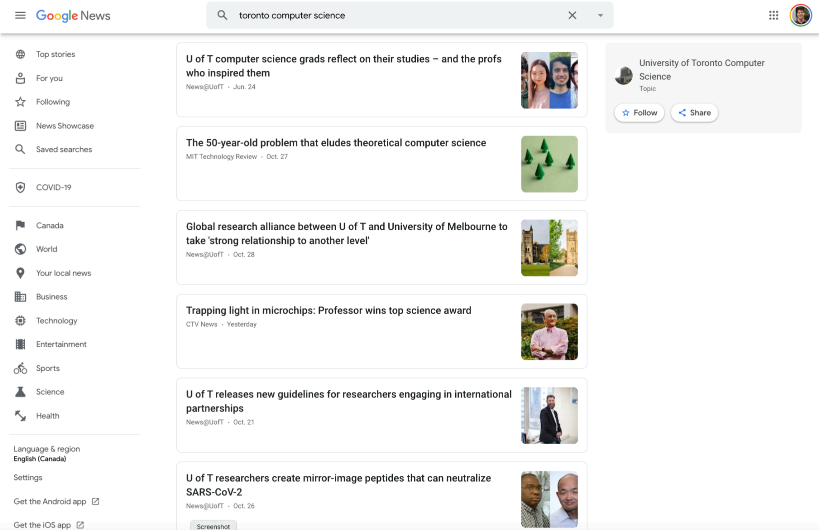
- If you were designing an ML algorithm to organize a user's social media feed, what other information might you use?

- As a supervised learning problem, what would be the inputs, and what would be the targets?

Challenge 1: Inferring User Preferences

Challenge 1: Inferring User Preferences

- Google News was an early example of training a model to predict clicks.



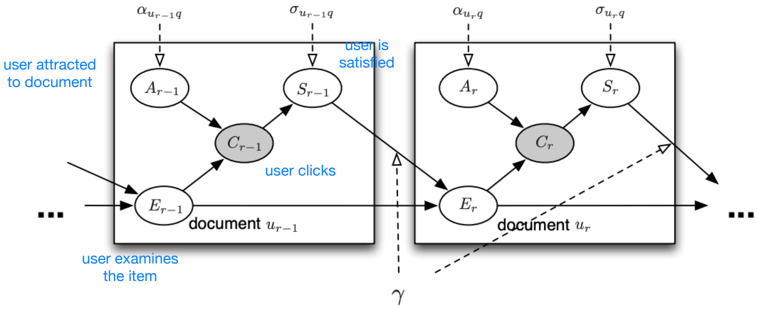
Challenge 1: Inferring User Preferences

- Why are clicks a useful signal?

- What are some problems with optimizing for clicks?

Challenge 1: Inferring User Preferences

- Here is a Bayesian network designed to model user behavior for a search engine.
  - We covered Bayes nets briefly when we discussed naıve Bayes.
- Nodes represent random variables, and edges represent direct influences. Shaded = observed.
- Want to infer user satisfaction (S).



Chuklin et al., “Click models for web search”

Challenge 1: Inferring User Preferences

- User preferences aren't just a matter of reactions to individual items, but also of the user's overall experience.
- Many web services optimize for a criterion called engagement.
  - User's frequency, intensity, or depth of interaction with a product over some time period
  - Not a technical term, but a business term, instantiated in different ways by different companies
  - E.g. Gmail: percentage of active users who visited the site on 5 or more days during the past week Rodden et al., "Measuring the user experience on a large scale"
  - E.g. Facebook: time spent on site, meaningful social interactions https://www.washingtonpost.com/technology/interactive/2021/how-facebook-algorithm-works/
- This is not directly optimized by an ML algorithm (as far as I know), but is used to evaluate changes to the system.
  - Sort of analogous to how logistic regression minimizes cross-entropy loss but you might tune hyperparameters based on accuracy.

Challenge 1: Inferring User Preferences

- The choice of what to optimize for can have ethical implications.
- The recently published Facebook Papers reveal a lot about unintended consequences of algorithm design
  - My aim isn't to pick on Facebook here. They found these harms and worked to fix them!
- Early years: optimizing for likes and clicks $\Rightarrow$ clickbait
- Optimizing for time spent reading/watching $\Rightarrow$ favored professional over organic content
- 2017: service changed to reward comments & emojis $\Rightarrow$ most successful political posts were the polarizing ones
  - Some political parties consciously shifted their messaging to be much more negative
  - Facebook eventually rolled back this change for health and politics

- https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215

Challenge 2: Bandit Feedback

Challenge 2: Bandit Feedback

- You only get information about user preferences for the posts you choose to show them. Therefore, the choices you make affect the data you get.
- This is closely related to the multi-armed bandit problem.

- You have a set of slot machine arms, and each arm i pays off \$1 with an unknown probability p_i.
- You are given T trials. You only find out the payoff for the arm that you tried. You want to maximize your total expected payoff.
- Showing the user a post = pulling an arm. Your metric (e.g. likes, clicks) = the payoff.

Challenge 2: Bandit Feedback

Here are the payoffs so far. Which arm should you pull next?

Arm 1: \$ \$ x \$ x \$ \$ x \$ \$

Arm 2: x x x x \$ \$ x x x

Arm 3: x \$

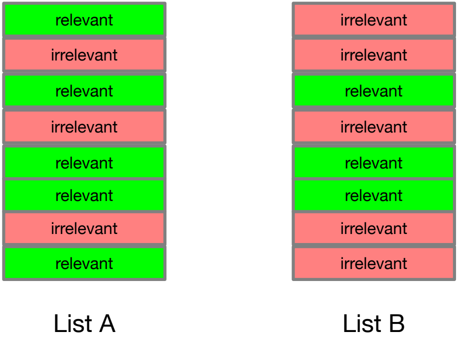
Challenge 2: Bandit Feedback

- Bandit problems are an important example of an exploration-exploitation tradeoff
  - "Exploitation": show the user a post you're confident they'll like
  - "Exploration": show the user a post they may or may not like so that you get information about their preferences

Challenge 3: Evaluating Structured Outputs

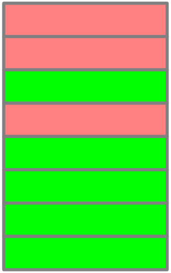
Challenge 3: Evaluating Structured Outputs

- Most of this class has focused on classification, where there is a natural metric to use (accuracy).
- In this case, we'd like to produce a feed (an ordered list of items). Problems where we want to predict a structured object are known as structured prediction.
- For now, assume that all items are either relevant or irrelevant.
- Which of the following lists is preferable?

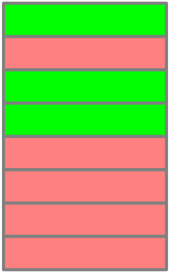


Challenge 3: Evaluating Structured Outputs

- One basic measure is precision: the fraction of items which are relevant.
- Which of the following lists is preferable?



List A



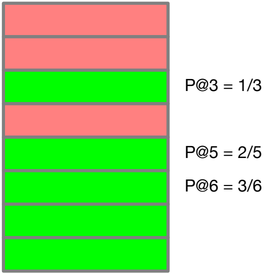
List B

Challenge 3: Evaluating Structured Outputs

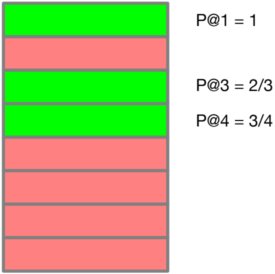
- Precision@K: Precision for the list up to the Kth item.
- Average Precision (AP): average of Precision@K, where K is taken as the indices of the first N relevant items.
  - Moving a relevant item from position 2 to position 1 is worth more points than moving it from position 8 to position 7.
- Mean Average Precision (MAP): mean of the AP over multiple queries.
- Note: in different application areas, there are different (but related) definitions of AP/MAP.

Challenge 3: Evaluating Structured Outputs

An example of calculating AP with $N = 3$.



$$\text{MAP} = \frac{1}{3}\left(\frac{1}{3} + \frac{2}{5} + \frac{3}{6}\right) \approx 0.41$$



$$\text{MAP} = \frac{1}{3}\left(1 + \frac{2}{3} + \frac{3}{4}\right) \approx 0.81$$

Challenge 3: Evaluating Structured Outputs

What other factors might you consider in evaluating a list of recommendations?

Towards Ethics

- We've been discussing challenges that arise when defining optimization objectives beyond the basic classification and regression settings.
- So far, we've focused on challenges of building a useful and engaging system.
- But what we choose to optimize for can have unintended consequences. The rest of the lecture focuses on thinking about optimization objectives from an ethical standpoint.

Acknowledgements

This module was created as part of an Embedded Ethics Education Initiative (E3I), a joint project between the Department of Computer Science[1] and the Schwartz Reisman Institute for Technology and Society[2], University of Toronto.

Instructional Team:

Roger Grosse, Steven Coyne, Emma McClure

Faculty Advisors:

Diane Horton[1], David Liu[1], and Sheila McIlraith[1,2]

Department of Computer Science
Schwartz Reisman Institute for Technology and Society
University of Toronto

Computer Science
UNIVERSITY OF TORONTO

UNIVERSITY OF TORONTO | SCHWARTZ REISMAN INSTITUTE FOR TECHNOLOGY AND SOCIETY