

2-4: Scoring and Ranking

Learning Objectives

- Understand several ways of computing and displaying predictions
- Understand how to rank items with sparse, time-shifting data
- Understand several points in the design space for prediction and recommendation, and some of their tradeoffs

Introduction

- Last 2 lectures:
 - how to collect data
 - what we present to users
- This lecture: how to do it
 - what predictions to show
 - how to rank

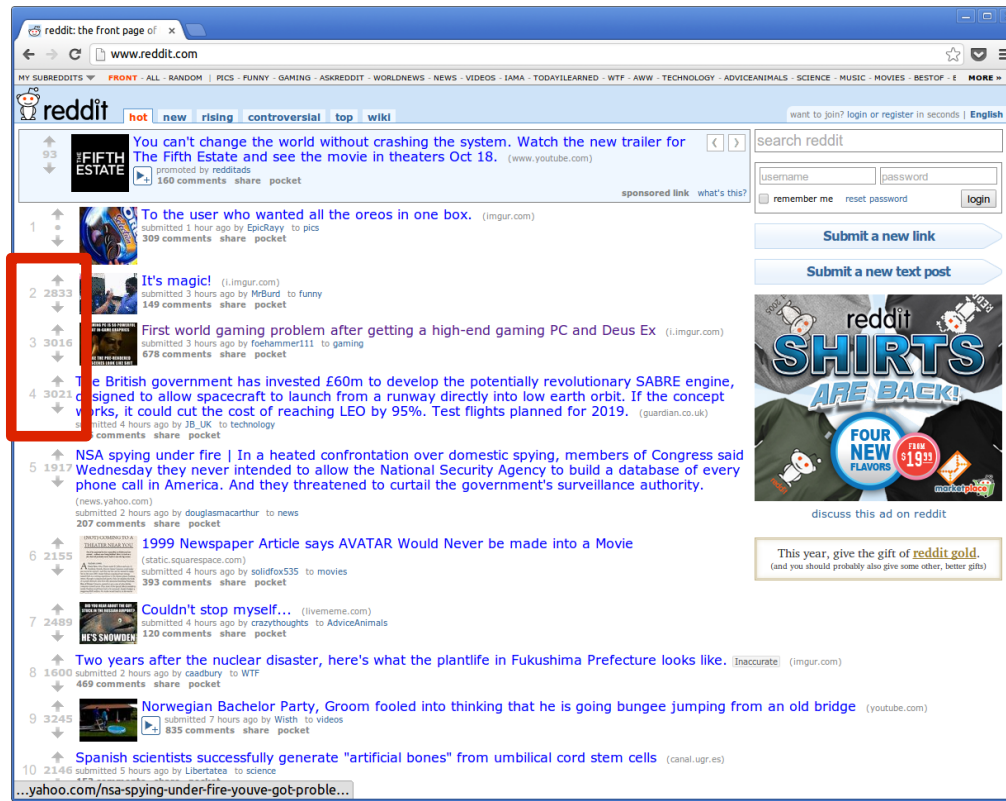
Overview

- Example
- Displaying Aggregate Preferences (*predict*)
- Ranking Items (*recommend*)

Overview

- Example
- Displaying Aggregate Preferences (*predict*)
- Ranking Items (*recommend*)

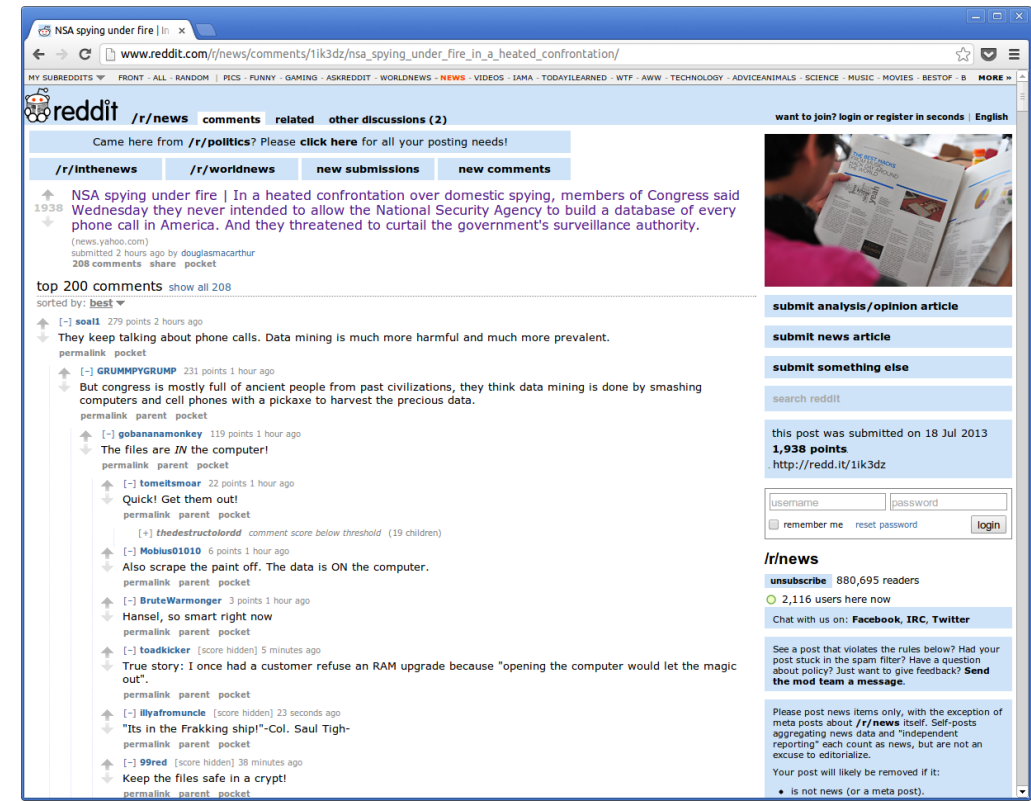
Introduction to Recommender Systems



Example - Reddit

- Social news aggregator
- Non-personalized news recommender
- Users vote on items to determine top item

Introduction to Recommender Systems



Overview

- Example
- **Displaying Aggregate Preferences (*predict*)**
- Ranking Items (*recommend*)

Introduction to Recommender Systems

Goal of Display

To help users decide to buy/read/view the item.

Introduction to Recommender Systems

Simple Display Approaches

- Average rating / upvote proportion
- Net upvotes / # of likes
- % ≥ 4 stars ('positive')
- Full distribution

Introduction to Recommender Systems

Simple Display Approaches (again)

- Average rating / upvote proportion
 - Of people who vote, do they like it?
 - Doesn't show popularity
- Net upvotes / # of likes
 - Shows popularity
 - No controversy
- % ≥ 4 stars ('positive')
- Full distribution

Introduction to Recommender Systems

– Complicated

Example: Amazon.com

Canon PowerShot A2300 16.0 MP Digital Camera with 5x Di Wide-Angle Lens with 720p HD Video Recording (Red)

by [Canon](#)

★★★★☆ (135 customer reviews)

4.2 out of 5 stars

5 star: (78)
4 star: (29)
3 star: (17)
2 star: (5)
1 star: (6)

[See all 135 reviews](#)

“It is easy to use and takes great pictures and video.”

Marc Jaffrey | 39 reviewers made a similar statement

“I highly recommend it for other point-and-shooters as well.”

Troildlocks | 12 reviewers made a similar statement

“It easily fits in my purse or pocket for easy carrying.”

Lois | 14 reviewers made a similar statement

Introduction to Recommender Systems

Overview

- Example
- Displaying Aggregate Preferences (*predict*)
- **Ranking Items (*recommend*)**

Introduction to Recommender Systems

Reddit

↑ 4 3021 The British government has invested £60m to develop the potentially revolutionary SABRE engine, designed to allow spacecraft to launch from a runway directly into low earth orbit. If the concept works, it could cut the cost of reaching LEO by 95%. Test flights planned for 2019. (guardian.co.uk)
submitted 4 hours ago by JB_UK to technology
645 comments share pocket

↑ 5 1917 NSA spying under fire | In a heated confrontation over domestic spying, members of Congress said Wednesday they never intended to allow the National Security Agency to build a database of every phone call in America. And they threatened to curtail the government's surveillance authority.
(news.yahoo.com)
submitted 2 hours ago by douglasmacarthur to news
207 comments share pocket

↑ 6 2155 1999 Newspaper Article says AVATAR Would Never be made into a Movie
(static.squarespace.com)
submitted 4 hours ago by solidfox535 to movies
393 comments share pocket

Introduction to Recommender Systems

Ranking

- What do you put at the top of Reddit?
- What is at the top of the e-Bay search list?
- You don't have to rank by prediction

Introduction to Recommender Systems

Why not rank by score?

- Too little data (one 5-star rating)
- Score may be multivariate (histogram)
- Domain or business considerations
 - Item is old
 - Item is ‘unfavored’

Introduction to Recommender Systems

Ranking Considerations

- Confidence
 - How confident are we that this item is good?
- Risk tolerance
 - High-risk, high-reward
 - Conservative recommendation
- Domain and business considerations
 - Age
 - System goals

Introduction to Recommender Systems

Damped means

- Problem: low confidence w/ few ratings
- Solution: assume that, without evidence, everything is average
- Ratings are evidence of non-averageness
- k controls strength of evidence required

$$\frac{\sum_u r_{ui} + k\mu}{n + k}$$

Introduction to Recommender Systems

Confidence Intervals

- From the reading: lower bound of statistical confidence interval (95%)
- Choice of bound affects risk/confidence
 - Lower bound is conservative: be sure it's good
 - Upper bound is risky: there's a chance of amazing
- Reddit uses Wilson interval (for binomial) to rank comments

Introduction to Recommender Systems

Domain Consideration: Time

- Reddit: old stories aren't interesting
 - even if they have many upvotes!
- eBay: items have short lifetimes

Introduction to Recommender Systems

Reddit algorithm (c. 2010)

$$\log_{10} \max(1, |U - D|) + \frac{\text{sign}(U - D) t_{\text{post}}}{45000}$$

- Log term applied to votes
 - decrease marginal value of later votes
- Time is seconds since Reddit epoch
- Buries items with negative votes
- Time vs. vote impact independent of age
- Scores news items, not comments

Introduction to Recommender Systems

Scoring news stories

- Hacker News

$$\frac{(U - D - 1)^\alpha}{(t_{\text{now}} - t_{\text{post}})^\gamma} \times P$$

- Net upvotes, polynomially decayed by age
- Old items scored mostly by vote
- Multiplied by item penalty terms
 - incorporate community goals into score

Introduction to Recommender Systems

Ranking Wrap-Up

- There are some theoretically grounded approaches (confidence interval, damping)
- Many sites use ad-hoc methods
- Most formulas have constants, will be highly service-dependent
- Can manipulate for ‘good’ or ‘evil’
- Build based on domain properties, goals

Introduction to Recommender Systems

Predict with sophisticated score?

- Theoretically a fine thing to do
- Be careful with transparency/scrutability
 - If you say ‘average rating’ for damped mean, and show ratings, users may be confused
 - Most important case (low ratings) also easiest to hand-verify

2-4: Scoring, Ranking, and Normalization

Conclusion

- Sparsity, inconsistency, temporal concerns make data messy
- Simple scoring doesn't necessarily match the domain or business
- There are good ways to deal with this (decay, time, penalties, damping)
- We'll see more normalizations later

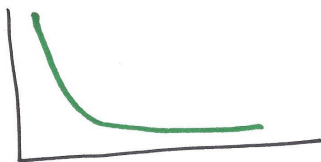
$$\frac{\sum_u r_{ui} + k\mu}{n + k}$$

strength required

$\mu=3$ 50% 50% $k=5$

net upvotes

$$\frac{(U-D-1)^{\alpha=0.8}}{(\tau_{\text{now}} - \tau_{\text{post}})^{\gamma=1.8}} \cdot \frac{P}{\text{Lage}}$$



$$\log_{10} \max(1, |u-d|) + \frac{\text{sign}(u-d) \tau_{\text{post}}}{45,000}$$

Pred.

$$P_{a,i} = \frac{\sum_{u=1}^n r_{u,i}}{n}$$
 ratings

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u)}{n}$$

$$P_{a,i} = \frac{\sum_{u=1}^n r_{u,i} \cdot w_{a,u}}{\sum_{u=1}^n w_{a,u}}$$

user a item i

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) * w_{a,u}}{\sum_{u=1}^n w_{a,u}}$$

similarity predictive-ness

$$w_{a,u} = \frac{\sum_{i=1}^m (r_{a,i} - \bar{r}_a)(r_{u,i} - \bar{r}_u)}{\sigma_a \sigma_u}$$

- m small?
- unary?

top-n?