Effects of User Similarity in Social Media

Ashton Anderson (Stanford)
Dan Huttenlocher (Cornell)
Jon Kleinberg (Cornell)
Jure Leskovec (Stanford)
User-to-user evaluations

Evaluations are ubiquitous on the web:

– People-items: most previous work
  • Collaborative Filtering
  • Recommendation Systems
  • E.g. Amazon

– People-people: our setting
Where does this occur on a large scale?

- **Wikipedia**: adminship elections
  - Support/Oppose (120k votes in English)
  - Four languages: English, German, French, Spanish

- **Stack Overflow**
  - Upvote/Downvote (7.5M votes)

- **Epinions**: Ratings of others’ product reviews (1-5 stars)
  - 5 = positive, 1-4 = negative
Goal

Understand what drives human evaluations
Overview of rest of the talk

1. What affects evaluations?
   – We will find that **status** and **similarity** are two fundamental forces

2. This will allow us to solve an interesting puzzle
   – Why are people so harsh on those who have around the same status as them?

3. Application: Ballot-Blind Prediction
   – We can accurately predict election outcomes without looking at the votes
Roadmap

1. What affects evaluations?
   - Status
   - Similarity
   - Status + Similarity

2. Solution to puzzle

3. Application: Ballot-blind prediction
Definitions

• **Status**
  – Level of recognition, merit, achievement in the community
  – Way to quantify: activity level
    • Wikipedia: # edits
    • Stack Overflow: # answers

• **User-user Similarity**
  – Overlapping topical interests of A and B
    • Wikipedia: cosine of articles edited
    • Stack Overflow: cosine of users evaluated
How does status affect the vote?

Natural hypothesis: \( \Pr[ + ] \sim f(S_B) \)

“Only attributes (e.g. status) of B matter”
How does status affect the vote?

Natural hypothesis: $\Pr[ + ] \sim f(S_B)$

“Only attributes (e.g. status) of B matter”

We find $\Pr[ + ] \sim f(S_A - S_B)$

Attributes of both evaluator and target are important

“Is B better than me?” is as important as “Is B good?”
Relative Status vs. P(+)

- Evaluator A evaluates target B
- \( P(+) \) as a function of \( \Delta = S_A - S_B \)?
- Intuitive hypothesis: monotonically decreases
How does similarity affect the vote?

Two natural (and opposite) hypotheses:

1. \( \uparrow \) similarity \( \Rightarrow \) \( \downarrow \) P(+)
   
   “The more similar you are, the better you can understand someone’s weaknesses”

2. \( \uparrow \) similarity \( \Rightarrow \) \( \uparrow \) P(+)
   
   “The more similar you are, the more you like the person”

Which one is it?
Second hypothesis is true:

↑ similarity  ⇔  ↑ P(+)

Large effect
How do similarity and status interact?

**Subtle relationship:** relative status matters a lot for low-similarity pairs, but doesn’t matter for high-similarity pairs.

Status is a proxy for more direct knowledge.

*Similarity controls the extent to which status is taken into consideration*
Who shows up to vote?

We find a selection effect in who gives the evaluations (on Wikipedia):
If \( S_A > S_B \), then A and B are highly similar.
What do we know so far?

1. Evaluations are diadic: \( \Pr[ + ] \sim f(S_A - S_B) \)

2. \( \uparrow \text{similarity} \implies \uparrow \text{P}(+) \)

3. Similarity controls how much status matters

4. In Wikipedia, high-status evaluators are similar to their targets
Roadmap

1. How user similarity affects evaluations

2. Solution to puzzle

3. Application: Ballot-blind prediction
Recall: Relative Status vs. P(+)

Intuitive hypothesis

Why?
Solution: similarity

Different mixture of P(+) vs. $S_A - S_B$ curves produces the mercy bounce

On Stack Overflow and Epinions, no selection effect and a different explanation
Roadmap

1. How user similarity affects evaluations

2. Solution to puzzle

3. Application: Ballot-blind prediction
Application: ballot-blind prediction

Task: Predict the outcome of a Wikipedia adminship election without looking at the votes

Why is this hard?
1. We can only look at the first 5 voters
2. We aren’t allowed to look at their votes

General theme: Guessing an audience’s opinion from a small fraction of the makeup of the audience
Features

1. Number of votes in each Δ-sim quadrant (Q)

2. Identity of first 5 voters (e.g. their previous voting history)

3. Simple summary statistics (SSS): target status, mean similarity, mean Δ

* Note now we are predicting on a per-instance basis, so it makes sense to use per-instance features
Our methods

Global method (**M1**):
\[
\Pr[E_i = 1] = P_i + d(\Delta_i, sim_i)
\]

Personal method (**M2**):
\[
\Pr[E_i = 1] = \alpha \ast P_i(\Delta_i, sim_i) + (1 - \alpha) \ast d(\Delta_i, sim_i)
\]

- **\(E_i\)**: \(i\)th evaluation
- **\(P_i\)**: voter \(i\)'s **positivity**: historical fraction of positive votes
- **\(d(\Delta_i, sim_i)\)**: global deviation from overall average vote fraction in \((\Delta_i, sim_i)\) quadrant
- **\(P_i(\Delta_i, sim_i)\)**: personal deviation
- **\(\alpha\)**: mixture parameter
Baselines and Gold Standard

- **Baselines:**
  - **B1:** Logistic regression with $\mathbf{Q} + \text{SSS}$
  - **B2:** $\Pr[E_i = 1] = P_i + \text{SSS}$

- **Gold Standard (GS)** cheats and looks at the votes
Results

English Wikipedia

German Wikipedia

Implicit feedback purely from audience composition
Summary
Thanks!

Questions?