Social Search
Networks of People and Search Engines

CS6200
Information Retrieval
Social Search

- Social search
  - Communities of users actively participating in the search process
  - Goes beyond classical search tasks
- Key differences
  - Users interact with the system
  - Users interact with other users either implicitly or explicitly
Web 2.0

- Social search includes, but is not limited to, the so-called social media sites
  - Collectively referred to as “Web 2.0” as opposed to the classical notion of the Web (“Web 1.0”)

- Social media sites
  - User generated content
  - Users can tag their own and other’s content
  - Users can share favorites, tags, etc., with others

- Examples (from the last 10 years):
  - Digg, Twitter, Flickr, YouTube, Del.icio.us, CiteULike, MySpace, Facebook, and LinkedIn
Social Search

- User tagging (i.e., manual indexing)
- Searching within communities
- Filtering and recommender systems
- Distributed search
  - Peer-to-peer (P2P, not covered here)
  - Metasearch (if there’s time)
User Tags and Manual Indexing

- Then: Library card catalogs
  - Indexing terms chosen with search in mind
  - Experts generate indexing terms
  - Terms are very high quality
  - Terms chosen from controlled vocabulary
- Now: Social media tagging
  - Tags not always chosen with search in mind
  - Users generate tags
  - Tags can be noisy or even incorrect
  - Tags chosen from *folksonomies*
**Title**  

**Author**  
Casellas, Nuria.

**Publisher**  

**Format**  
E-Book

**Physical Descrip.**  
1 online resource (xxii, 297 p.)

**Series**  
( Law, governance and technology series ; v.3 )

**Bibliography**  
Includes bibliographical references and index.

**Subject(s)**  
Legal research -- Data processing.  
Ontologies (Information retrieval)  
Semantic computing.

**Genre**  
Electronic books.

**Other Author(s)**  
SpringerLink (Online service)

**Related Entry**  
Springer e-books

**Series Added Entry**  
( Law, governance and technology series ; v.3 )

**ISBN**  
9789400714977 (electronic bk.)  
9400714971 (electronic bk.)

**Location(s)**  
SC Neilson Library / SC Internet / Online Resource
Social Tagging

- Flickr screenshot with tags: cherry blossom 16, cherry blossom
- Photo taken on April 17, 2012
- Viewed 6 times
- Belongs to Honeybuy2012's photostream (439)
- License: All Rights Reserved
- Privacy: This photo is visible to everyone
Social Tagging

President Obama visits the Cherry Blossom Festival? Well, not really — his motorcade passed by the Tidal Basin en route to Ft. McNair for his big Libya speech. A few lucky tourists got a chance to see it whizz by.

Comments and faves

sarahrenlca added this photo to their favorites. (13 months ago)

Tags

washington dc • district of columbia • 15th street sw • tidal basin • 2011 cherry blossom festival • bureau of printing and engraving • department of the treasury • president barack obama • motorcade • presidential limousine • aka the beast • dcist • POTUS • cadillac
Social Tagging
Some Categories of Tags

- Content-based: cherry blossoms, car
- Context-based: Washington, DC
- Attributes: Nikon, B&W
- Subjective: delicious, awesome
- Organizational: to_read
Searching Tags

• Searching user tags is challenging
  ✤ Most items have only a few tags
  ✤ Tags are very short

• Boolean, probabilistic, vector space, and language modeling will fail if use naïvely

• Must overcome the vocabulary mismatch problem between the query and tags
Tag Expansion

• Can overcome vocabulary mismatch problem by expanding tag representation with external knowledge

• Possible external sources
  * Thesaurus
  * Web search results
  * Query logs

• After tags have been expanded, can use standard retrieval models
Tag Expansion

Age of Aquariums - Tropical Fish
Huge educational aquarium site for tropical fish hobbyists, promoting responsible fish keeping internationally since 1997.

The Krib (Aquaria and Tropical Fish)
This site contains information about tropical fish aquariums, including archived usenet postings and e-mail discussions, along with new ...

...

Keeping Tropical Fish and Goldfish in Aquariums, Fish Bowls, and ...
Keeping Tropical Fish and Goldfish in Aquariums, Fish Bowls, and Ponds at AquariumFish.net.
Searching Tags

• Even with tag expansion, searching tags is challenging

• Tags are inherently noisy and incorrect

• Many items may not even be tagged!

• Typically easier to find popular items with many tags than less popular items with few/no tags
Inferring Missing Tags

• How can we automatically tag items with few or no tags?

• Uses of inferred tags
  * Improved tag search
  * Automatic tag suggestion
Inferring Tags

- **TF.IDF**
  - Suggest tags that have a high TF.IDF weight in the item
  - Only works for textual items

- **Classification**
  - Train binary classifier for each tag
  - Performs well for popular tags, but not as well for rare tags

- **Maximal marginal relevance**
  - Finds tags that are relevant to the item and novel with respect to existing tags

\[
M M R(t; T_i) = \left( \lambda S i m_{item}(t, i) - (1 - \lambda) \max_{t \in T_i} S i m_{tag}(t_i, t) \right)
\]
Browsing & Tag Clouds

- Search is useful for finding items of interest
- Browsing is more useful for exploring collections of tagged items
- Various ways to visualize collections of tags
  * Tag lists
  * Tag clouds
  * Alphabetical order
  * Grouped by category
  * Formatted/sorted according to popularity
Tag Clouds

animals architecture art australia autumn baby birth barcelona beach berlin birthday black blackandwhite blue california cameraphone canadacan

car cat chicago china christmas church city clouds color concert day dog england europe family festival film florida flower flowers food france friends fun garden germany girl graffiti green halloween hawaii holiday home house india ireland italy japan july kids lake landscape light live london macro me mexico music nature new newyork night nikon nyc ocean paris park party people portrait red river rock sanfrancisco scotland sea seattle show sky snow spain spring street summer sunset taiwan texas thailand tokyo toronto travel tree trees trip uk usa vacation washington water wedding
Searching within Communities
Conversational Structure
Conversational Structure

5. Re: Chihuly exhibit at MFA - simply fantastic!
29 June 2011, 4:57

Mahj
Destination Expert for Boston
Been twice so far, may go a third time. Quite a boon for the MFA, but do make sure you wander through the other exhibits as well.

6. Re: Chihuly exhibit at MFA - simply fantastic!
29 June 2011, 6:34

bellman3
Massachusetts
It really is very wonderful. We didn’t know that photography was permitted, so I definitely want to go again to take pics. But even if I don’t have photos, I don’t think my memories of the pieces will fade anytime soon.

7. Re: Chihuly exhibit at MFA - simply fantastic!
29 June 2011, 8:43

sojoh
Paris, France
Mahj, I agree that the other exhibits at the MFA are well worth a visit too. It was a real pleasure to see the Impressionist paintings and to see a bit of the new Art of Americas wing. I’m already planning to return to the MFA during my next visit to Boston.
Conversational Structure

5. Re: Chihuly exhibit at MFA - simply fantastic!
29 June 2011, 4:57

Destination Expert
for Boston

Been twice so far, may go a third time. Quite a boon for the MFA, but do make sure you wander through the other exhibits as well.

6. Re: Chihuly exhibit at MFA - simply fantastic!
29 June 2011, 6:34

It really is very wonderful. We didn’t know that photography was permitted, so I definitely want to go again to take pics. But even if I don’t have photos, I don’t think my memories of the pieces will fade anytime soon.

7. Re: Chihuly exhibit at MFA - simply fantastic!
29 June 2011, 8:43

Mahj, I agree that the other exhibits at the MFA are well worth a visit too. It was a real pleasure to see the Impressionist paintings and to see a bit of the new Art of Americas wing. I’m already planning to return to the MFA during my next visit to Boston.

Seo, Croft, Smith /IR 2011
Conversational Structure

Seo, Croft, Smith / IR 2011
Retrieving Social Data

NDCG@10

Post  Thread  Pair  Dialogue  Dial.+Thread

Seo, Croft, Smith IR 2011
Finding Communities

• How can we find users with common interests?
• How can we find documents on a common topic?
• Graph clustering
• Hypertext Induced Topic Search (HITS)
  ✤ Nodes may be Hubs or Authorities
  ✤ Iterative solution
Algorithm 3 HITS

1: procedure HITS($G = (V, E)$, $K$)
2:     $A_0(p) \leftarrow 1 \forall p \in V$
3:     $H_0(p) \leftarrow 1 \forall p \in V$
4:     for $i = 1$ to $K$
5:         $A_i(p) \leftarrow 0 \forall p \in V$
6:         $H_i(p) \leftarrow 0 \forall p \in V$
7:         $Z_A \leftarrow 0$
8:         $Z_H \leftarrow 0$
9:         for $p \in V$
10:             for $q \in V$
11:                 if $(p, q) \in E$
12:                     $H_i(p) \leftarrow H_i(p) + A_{i-1}(q)$
13:                     $Z_H \leftarrow Z_H + A_{i-1}(q)$
14:                 end if
15:             if $(q, p) \in E$
16:                 $A_i(p) \leftarrow A_i(p) + H_{i-1}(q)$
17:                 $Z_A \leftarrow Z_A + H_{i-1}(q)$
18:             end if
19:         end for
20:     end for
21:     for $p \in V$
22:         $A_i(p) \leftarrow A_i(p) \frac{Z_A}{Z_H}$
23:         $H_i(p) \leftarrow H_i(p) \frac{Z_H}{Z_H}$
24:     end for
25: end procedure
26: return $A_K$, $H_K$
27: end procedure
HITS Iteration
Clustering Nodes

Node:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Vector:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
EM Clustering (for k clusters)
EM Clustering (for k clusters)

- EM algorithm
  - Viterbi version – called “k-means clustering”
  - Full EM version – called “Gaussian mixtures”
EM Clustering (for k clusters)

- EM algorithm
  - Viterbi version – called “k-means clustering”
  - Full EM version – called “Gaussian mixtures”
EM Clustering (for k clusters)

- EM algorithm
  - Viterbi version – called “k-means clustering”
  - Full EM version – called “Gaussian mixtures”

- **Expectation step**: Use current parameters (and observations) to reconstruct hidden structure
EM Clustering (for k clusters)

- EM algorithm
  - Viterbi version – called “k-means clustering”
  - Full EM version – called “Gaussian mixtures”

- **Expectation step:** Use current parameters (and observations) to reconstruct hidden structure
- **Maximization step:** Use that hidden structure (and observations) to reestimate parameters
EM Clustering (for k clusters)

- EM algorithm
  - Viterbi version – called “k-means clustering”
  - Full EM version – called “Gaussian mixtures”

- **Expectation step:** Use current parameters (and observations) to reconstruct hidden structure

- **Maximization step:** Use that hidden structure (and observations) to reestimate parameters
EM Clustering (for k clusters)

- EM algorithm
  - Viterbi version – called “k-means clustering”
  - Full EM version – called “Gaussian mixtures”

- **Expectation step:** Use current parameters (and observations) to reconstruct hidden structure

- **Maximization step:** Use that hidden structure (and observations) to reestimate parameters

- **Parameters:** k points representing cluster centers
EM Clustering (for k clusters)

- **EM algorithm**
  - Viterbi version – called “k-means clustering”
  - Full EM version – called “Gaussian mixtures”

- **Expectation step:** Use current parameters (and observations) to reconstruct hidden structure
- **Maximization step:** Use that hidden structure (and observations) to reestimate parameters

- **Parameters:** k points representing cluster centers
- **Hidden structure:** for each data point (word type), which center generated it?
Searching with Communities

• Identify communities
  ✤ Graph clustering
  ✤ Hypertext induced topic search

• Exploiting community knowledge
  ✤ Authorities in HITS graph
  ✤ Explicit question answering and feedback
  ✤ Bonus: Training data for retrieval models
Community QA

Search Results

What are some pagerank alternatives?
... (for example, pagerank knows who’s linking(edges) to whom)

getting pagerank
... how can I retrieve pagerank of any page indexed by google? Has google any api or page for this?

pagerank implementation in java
... I am looking for a java implementation of the pagerank algorithm.
What are some pagerank alternatives?

As well as HITS [as suggested by @larsmans], there is also SALSA, which is considered more "stable" from HITS [and thus is less vulnerable to be affected by spammers].

You are also encouraged to have a look at this survey or ranking algorithms.

Twitter uses SALSA in their user recommendation technique. – Steve Jan 10 at 22:15
Community QA

• Pros
  ✤ Answers to complex information needs
  ✤ Compare multiple opinions
  ✤ Feedback, interaction with others

• Cons
  ✤ Latency
  ✤ All the drawbacks of human interaction
Community QA

What part of Mexico gets the most tropical storms?
How do you pronounce the french words, coeur and miel?
GED test?
Why do I have to pay this fine?
What is Schrödinger’s cat?
What’s this song?
Hi...can u ppl tell me sumthing abt death dreams??
What are the engagement and wedding traditions in Egypt?
Fun things to do in LA?
What lessons from the Tao Te Ching do you apply to your everyday life?
Foci of a hyperbola?
What should I do today?
Why was iTunes deleted from my computer?
Heather Locklear?
Do people in the Australian Defense Force (RAAF) pay less tax than civilians?
What's a psp xmb?
If C(-3, y) and D(1, 7) lie upon a line whose slope is 2, find the value of y.?
Why does love make us so irrational?
Am I in love?
What are some technologies that are revolutionizing business?
Community QA

- What part of Mexico gets the most tropical storms?
- How do you pronounce the french words, coeur and miel?
- GED test?
- Why do I have to pay this fine?
- What is Schrödinger’s cat?
- What’s this song?
- Hi... can u ppl tell me sumthing abt death dreams??
- What are the engagement and wedding traditions in Egypt?
- Fun things to do in LA?
- What lessons from the Tao Te Ching do you apply to your everyday life?
- Foci of a hyperbola?
- What should I do today?
- Why was iTunes deleted from my computer?
- Heather Locklear?
- Do people in the Australian Defense Force (RAAF) pay less tax than civilians?
- What’s a psp xmb?
- If C(-3, y) and D(1, 7) lie upon a line whose slope is 2, find the value of y.?
- Why does love make us so irrational?
- Am I in love?
- What are some technologies that are revolutionizing business?
IR in Communities

• How to match queries to community QA (or forums, or other social media)?
  * Match query to questions, answers, both?
  * Generally more effective to match questions

• Questions, and other posts, are short

• More problems with *vocabulary mismatch*
Retrieval as Translation
Retrieval as Translation

Naive translation model

\[ P(Q|A) = \prod_{w \in Q} \sum_{t \in V} P(w|t) P(t|A) \]
Retrieval as Translation

Naive translation model

\[ P(Q|A) = \prod_{w \in Q} \sum_{t \in \mathcal{V}} P(w|t) P(t|A) \]

Interpolation with language model

\[ P(Q|A) = \prod_{w \in Q} \frac{(1 - \beta) f_{w,A} + \beta \sum_{t \in \mathcal{V}} P(w|t) f_{t,A} + \mu \frac{c_w}{|\mathcal{C}|}}{|A| + \mu} \]
Retrieval as Translation

Naive translation model

\[ P(Q|A) = \prod_{w \in Q} \sum_{t \in V} P(w|t) P(t|A) \]

Interpolation with language model

\[ P(Q|A) = \prod_{w \in Q} \frac{(1 - \beta)f_{w,A} + \beta \sum_{t \in V} P(w|t)f_{t,A} + \mu \frac{e_w}{c_w}}{|A| + \mu} \]

Example “translations”

<table>
<thead>
<tr>
<th>everest</th>
<th>xp</th>
<th>search</th>
</tr>
</thead>
<tbody>
<tr>
<td>mountain</td>
<td>window</td>
<td>google</td>
</tr>
<tr>
<td>tallest</td>
<td>install</td>
<td>information</td>
</tr>
<tr>
<td>29,035 ft</td>
<td>drive</td>
<td>internet</td>
</tr>
<tr>
<td>highest mt</td>
<td>computer</td>
<td>website</td>
</tr>
<tr>
<td>feet</td>
<td>version</td>
<td>web</td>
</tr>
<tr>
<td>mount</td>
<td>click</td>
<td>list</td>
</tr>
<tr>
<td>program</td>
<td>pc</td>
<td>free</td>
</tr>
<tr>
<td>microsoft</td>
<td></td>
<td>info</td>
</tr>
<tr>
<td></td>
<td></td>
<td>page</td>
</tr>
</tbody>
</table>
Retrieval as Translation

Naive translation model

\[ P(Q|A) = \prod_{w \in Q} \sum_{t \in V} P(w|t) P(t|A) \]

Interpolation with language model

\[ P(Q|A) = \prod_{w \in Q} \frac{(1 - \beta) f_{w,A} + \beta \sum_{t \in V} P(w|t)f_{t,A} + \mu c_w}{|A| + \mu} \]

Example “translations”

<table>
<thead>
<tr>
<th>everest</th>
<th>xp</th>
<th>search</th>
</tr>
</thead>
<tbody>
<tr>
<td>mountain</td>
<td>window</td>
<td>google</td>
</tr>
<tr>
<td>tallest</td>
<td>install</td>
<td>information</td>
</tr>
<tr>
<td>29,035</td>
<td>drive</td>
<td>internet</td>
</tr>
<tr>
<td>highest</td>
<td>computer</td>
<td>website</td>
</tr>
<tr>
<td>mt</td>
<td>version</td>
<td>web</td>
</tr>
<tr>
<td>ft</td>
<td>click</td>
<td>list</td>
</tr>
<tr>
<td>measure</td>
<td>pc</td>
<td>free</td>
</tr>
<tr>
<td>feet</td>
<td>program</td>
<td>info</td>
</tr>
<tr>
<td>mount</td>
<td>microsoft</td>
<td>page</td>
</tr>
</tbody>
</table>

How to estimate \( P(w|t) \)?

Q/A pairs!
Collaborative Search

The community searches together

Co-located Collaborative Searching

Remote Collaborative Searching

Shared queries, results, relevance judgments, etc.
Filtering and Recommending
Filtering and Recommending

• User profiles the fundamental data structure
  ✤ Profiles can be static or dynamic
  ✤ Query features + user/social features

• Filtering
  ✤ Canonical case: query an endless stream

• Recommending (a.k.a. “collaborative filtering”)
  ✤ Jointly infer relevance from lots of profiles
Static Filtering
Adaptive Filtering
Filtering Models
Filtering Models

Profile model

\[ P(w|P) = \frac{(1 - \lambda)}{\sum_{i=1}^{K} \alpha_i} \sum_{i=1}^{K} \alpha_i \frac{f_{w,T_i}}{|T_i|} + \lambda \frac{c_w}{|C|} \]
Filtering Models

Profile model

\[
P(w|P) = \frac{(1 - \lambda) \sum_{i=1}^{K} \alpha_i f_{w,T_i}}{\sum_{i=1}^{K} \alpha_i |T_i|} + \lambda \frac{c_w}{|C|}
\]

Document model

\[
P(w|D) = (1 - \lambda) \frac{f_{w,D}}{|D|} + \lambda \frac{c_w}{|C|}
\]
Filtering Models

Profile model

\[ P(w|P) = \frac{(1 - \lambda)}{\sum_{i=1}^{K} \alpha_i} \sum_{i=1}^{K} \alpha_i \frac{f_{w,T_i}}{|T_i|} + \lambda \frac{c_w}{|C|} \]

Document model

\[ P(w|D) = (1 - \lambda) \frac{f_{w,D}}{|D|} + \lambda \frac{c_w}{|C|} \]

Relevance (adaptive) model

\[ P(w|P) = \frac{1}{|Rel|} \sum_{D_i \in Rel} \sum_{D \in C} P(w|D)P(D_i|D) \approx \frac{1}{|Rel|} \sum_{D_i \in Rel} P(w|D_i) \]
Filtering Models

Profile model

\[ P(w|P) = \frac{(1 - \lambda)}{\sum_{i=1}^{K} \alpha_i \frac{f_{w,T_i}}{|T_i|}} + \lambda \frac{c_w}{|C|} \]

Relevance (adaptive) model

\[ P(w|P) = \frac{1}{|\text{Rel}|} \sum_{D_i \in \text{Rel}} \sum_{D \in C} P(w|D) P(D_i|D) \]
\[ \approx \frac{1}{|\text{Rel}|} \sum_{D_i \in \text{Rel}} P(w|D_i) \]

Document model

\[ P(w|D) = (1 - \lambda) \frac{f_{w,D}}{|D|} + \lambda \frac{c_w}{|C|} \]

Kullback-Leibler divergence

\[ -KL(P||D) = \sum_{w \in V} P(w|P) \log P(w|D) - \sum_{w \in V} P(w|P) \log P(w|P) \]
Filtering Models

Profile model

\[
P(w|P) = \frac{(1 - \lambda)}{\sum_{i=1}^{K} \alpha_i} \sum_{i=1}^{K} \alpha_i \frac{f_{w,T_i}}{|T_i|} + \lambda \frac{c_w}{|C|}
\]

Relevance (adaptive) model

\[
P(w|P) = \frac{1}{|Rel|} \sum_{D_i \in Rel} \sum_{D \in C} P(w|D) P(D_i|D)
\approx \frac{1}{|Rel|} \sum_{D_i \in Rel} P(w|D_i)
\]

Document model

\[
P(w|D) = (1 - \lambda) \frac{f_{w,D}}{|D|} + \lambda \frac{c_w}{|C|}
\]

Kullback-Leibler divergence

\[
-KL(P||D) = \sum_{w \in V} P(w|P) \log P(w|D) - \sum_{w \in V} P(w|P) \log P(w|P)
\]

<table>
<thead>
<tr>
<th>Model</th>
<th>Profile Representation</th>
<th>Profile Updating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boolean</td>
<td>Boolean Expression</td>
<td>N/A</td>
</tr>
<tr>
<td>Vector Space</td>
<td>Vector</td>
<td>Rocchio</td>
</tr>
<tr>
<td>Language Modeling</td>
<td>Probability Distribution</td>
<td>Relevance Modeling</td>
</tr>
<tr>
<td>Classification</td>
<td>Model Parameters</td>
<td>Online Learning</td>
</tr>
</tbody>
</table>
Scaling and Evaluation

- Treat profiles as “documents” and index them
- Treat incoming documents as queries
- Tradeoffs vary by application

<table>
<thead>
<tr>
<th></th>
<th>Relevant</th>
<th>Non-relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrieved</td>
<td>TP</td>
<td>FP</td>
</tr>
<tr>
<td>Not retrieved</td>
<td>FN</td>
<td>TN</td>
</tr>
</tbody>
</table>

\[ U = \alpha \cdot TP + \beta \cdot TN + \delta \cdot FP + \gamma \cdot FN \]
Recommending
Imagine there's only one document/product
Imagine there's only one document/product.

Similar users probably make similar judgments and have similar tastes.
Imagine there's only one document/product

User without mainstream tastes?

Similar users probably make similar judgments and have similar tastes.
Rating by Distance

- Cluster users based on overall rating similarity (or other features)

\[
\hat{r}_u(i) = \frac{1}{|\text{Cluster}(u)|} \sum_{u' \in \text{Cluster}(u)} r_{u'}(i)
\]

- Assign unknown ratings the cluster average

\[
\hat{r}_u(i) = \overline{r}_u + \frac{1}{\sum_{u' \in \mathcal{N}(u)} \text{sim}(u, u')} \sum_{u' \in \mathcal{N}(u)} \text{sim}(u, u')(r_{u'}(i) - \overline{r}_w)
\]

- Average nearest neighbors
Evaluating Recommenders

- Exact-match accuracy (usually too harsh)
- Absolute error

\[ ABS = \frac{1}{|U||I|} \sum_{u \in U} \sum_{i \in I} |\hat{r}_{u}(i) - r_{u}(i)| \]

- Mean squared error

\[ MSE = \frac{1}{|U||I|} \sum_{u \in U} \sum_{i \in I} (\hat{r}_{u}(i) - r_{u}(i))^2 \]

- Other task-dependent measures