2005

Recommender Systems Research

Saverio Perugini
University of Dayton, sperugini1@udayton.edu

Follow this and additional works at: http://ecommons.udayton.edu/cps_fac_pub

Part of the Databases and Information Systems Commons, Numerical Analysis and Scientific Computing Commons, and the Other Computer Sciences Commons

eCommons Citation
http://ecommons.udayton.edu/cps_fac_pub/32

This Conference Paper is brought to you for free and open access by the Department of Computer Science at eCommons. It has been accepted for inclusion in Computer Science Faculty Publications by an authorized administrator of eCommons. For more information, please contact frice1@udayton.edu, mschlangen1@udayton.edu.
Recommender Systems Research

Saverio Perugini
Department of Computer Science
University of Dayton
“What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.”

Herbert A. Simon
Recommender Systems

- Select a subset of items based on user preferences
- Underlying algorithms range from simple keyword matching to sophisticated mining of user profiles
- Examples: top-N lists, book and movie recommenders: Amazon.com
- Reduce information overload
- Retain customers
- Increase revenue
- Now believed to be critical to sustaining the Internet economy
Four Main Dimensions

How is the recommender system

1. modeled and designed
   • are recommendations content-based or collaborative?

2. targeted
   • to an individual, group, or topic?

3. built

4. maintained
   • online vs. offline
Content-based Filtering

‘Since you liked *The Little Lisper*,
you may be interested in *The Little Schemer.*’

‘Since you liked *Pride and Prejudice*,
you also might like *Sense and Sensibility.*’
Collaborative-filtering

Linus and Lucy like *Sleepless in Seattle*. Linus likes *You’ve Got Mail*.

Lucy also might like *You’ve Got Mail*. 
Four Main Dimensions

How is the recommender system

1. modeled and designed
   - are recommendations content-based or collaborative?
2. targeted
   - to an individual, group, or topic?
3. built
4. maintained
   - online vs. offline

What about the inherently social aspect of recommendation?
‘Brick and Mortar’ Setting

[Linus and Lucy are at Bombay Café, an Indian Restaurant.]

1  Linus:  The menu looks enticing.
2  Linus:  Since you are a returning patron, what do you recommend?
3  Lucy:  Well, since you like spicy dishes, and you’re not a vegetarian, you’ll enjoy the Chicken Vindaloo.
4  Linus:  Alright, I’ll try that.

A mutually-reinforcing dynamic ensues:

• Lucy leverages her knowledge of Linus’ interests into the process of recommendation.

• Linus harnesses his knowledge of Lucy’s reputation to evaluate the recommendation.
Inherent Social Aspect

• Recommender systems attempt to emulate and automate this natural social process.

• Predictive utility relies on its representation of the recipient.

• Recommender systems involve *user modeling*.

• User models can be constructed by
  
  – explicitly soliciting feedback
    
    * e.g., asking users to rate products or services
  
  – gleaning implicit declarations of interest
    
    * e.g., through monitoring usage
A Connection-centric View

User Modeling

Explicit

Social Network

Implicit

connections

keywords, surveys, reviews, feedback

formation

user-generated data

DB

DB

communication logs

documents, Usenet msgs

the Web

composite representation of user
(e.g., ratings or profile)

discovery

Recommender Systems Research MAICS’05
**Shifts in IS Research**

<table>
<thead>
<tr>
<th>Concept</th>
<th>Modeling Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information retrieval</td>
<td>(terms (\times) documents)</td>
</tr>
<tr>
<td></td>
<td>↓</td>
</tr>
<tr>
<td>Information filtering</td>
<td>(features (\times) documents)</td>
</tr>
<tr>
<td></td>
<td>↓</td>
</tr>
<tr>
<td>Content-based filtering</td>
<td>(features (\times) artifacts)</td>
</tr>
<tr>
<td></td>
<td>↓</td>
</tr>
<tr>
<td>Collaborative filtering</td>
<td>(people (\times) documents)</td>
</tr>
<tr>
<td></td>
<td>↓</td>
</tr>
<tr>
<td>Recommender systems</td>
<td>(people (\times) artifacts)</td>
</tr>
</tbody>
</table>
User Modeling Methodology for CF RSs

- user reluctance to rate items (compounded by volume & concern of privacy)
  - sparse modeling matrix (cold-start)
  - explicit + implicit user modeling (*exploration*)
  - representation of user (ratings, profiles) as basis for connection
  - deliver recommendations & create connections (*exploitation*)
Explicit User Modeling

• What kind?
  – quantitative (e.g., ratings)
  – qualitative (e.g., reviews)

• What makes it tough?
  – voluminous (and ephemeral) domains (e.g., news)
  – reluctance to evaluate artifacts
  – free-riders
  – cold-start: new user or new item
  – ‘banana’ problem (and converse)
  – users with unusual or highly specific tastes
  – users with similar interests who have rated different artifacts
  – effusivity of ratings
Explicit User Modeling (contd)

- Possible solutions?
  - pay-per-use model, subscription services
  - minimum rating constraints
  - incentives
  - default votes
  - agents to rate every artifact
  - user interface approaches
  - rate clusters of items
  - hybrid approaches (collaborative and content-based)
  - use indirection

- Representative projects
  - GroupLens (Pearson’s $r$)
  - Fab (hybrid)
Implicit User Modeling

• Traditional approaches
  – PHOAKS (USENET News)
  – Siteseer (bookmarks)

• Link analysis and cyber-communities
  – Social networks
    * Discovering shared interests
    * Referral Web

• Mining and exploiting structure
  – Jumping connections
  – HITS: Hubs and authorities

• Small-world networks
What are implicit declarations of interest?

- Clickstream data, (web) access logs, ‘footprints’
- Time spent on a product page
- UI events: scrolling, highlighting
- Transaction data, shopping carts
- Hyperlinks
- Bookmarks
PHOAKS

recommenders

p1
p2
p3

recipients

(3)
URL

p4
p5
p6
Link Analysis and Cyber-communities

- Discovering shared interests
  - Used e-mail logs to mine connections
  - Closeness

\[
\text{InterestDistance} (n_1, n_2) = \frac{|(C(n_1) \cup C(n_2)) - (C(n_1) \cap C(n_2))|}{|(C(n_1) \cup C(n_2))|}
\]

- Referral Web
  - Used close proximity of names in webpages
  - Queries:
    * Referral chains: ‘What is my relationship to Marvin Minsky?’
    * Search for experts: ‘What colleagues of mine, or colleagues of colleagues of mine know about simulated annealing?’
    * Proximity search: ‘List documents on the topic annealing by people close to Scott Kirkpatrick.’
Mining and Exploiting Structure: Theme

- **Mine Structure**: Why does the structure arise in the first place?
- **Model Structure**: Exploit Structure
Mining and Exploiting Structure (contd)

- Affiliation networks vs. social networks
  - actor-movie collaboration graph
  - author-paper collaboration graph

- What structure can be mined?
  - degree distribution
  - connectivity

- Examples:
  - Jumping Connections
  - HITS: Hubs and authorities
Jumping Connections

\[ J \]

(a) p1
   p2
   p3
   p4
   p5
   m1
   m2
   m3
   m4

(b) p1
    p2
    p3
    p4
    p5

(c) m1
    m2
    m3
    m4

Recommender Systems Research

MAICS’05
HITS: Hubs and Authorities
Small-world Networks

Regular Network

Small-World Network

Random Network

Increasing randomness

\[ p = 0 \quad \text{Increasing randomness} \quad p = 1 \]
# Discovering Social Networks

<table>
<thead>
<tr>
<th>Concept</th>
<th>Implicit declaration of interest</th>
<th>Algorithm or System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional Approaches to Implicit User Modeling</td>
<td>URLs in Usenet news bookmarks</td>
<td>PHOAKS&lt;br&gt;Siteseeer</td>
</tr>
<tr>
<td>Link Analysis and Cyber-Communities</td>
<td>e-mail logs&lt;br&gt;web documents</td>
<td>Discovering Shared Interests&lt;br&gt;Referral Web</td>
</tr>
<tr>
<td>Mining and Exploiting Structure</td>
<td>movie ratings datasets&lt;br&gt;hits-buffs, half bow-tie&lt;br&gt;web link topology&lt;br&gt;hubs and authorities&lt;br&gt;bow-tie</td>
<td>Jumping Connections&lt;br&gt;PageRank (Google)&lt;br&gt;HITS (CLEVER)</td>
</tr>
<tr>
<td>Small-world Networks</td>
<td>actor collaborations&lt;br&gt;author collaborations&lt;br&gt;infectious disease&lt;br&gt;the web</td>
<td>Oracle of Bacon&lt;br&gt;DBLP</td>
</tr>
</tbody>
</table>
Take away

• Purely structural information can be very instructive.
• These properties are found in nature (self-generating and self-organizing systems) and not merely an artifact of an idealized world.
• In what ways can we exploit these properties for recommendation?
Broadening Issues

• Evaluation
  – Functional vs. human-oriented evaluations
  – Is there something in between?

• Targeting
  – Answers the question ‘for whom are we building this system?’

• Privacy and trust
  – Broader than one user and one system
  – Concept of a weak-tie

• Shilling
  – Involves inundating the system with data intended to coerce it to artificially recommend the perpetrator’s products more often than those of a competitor.
  – Algorithms to detect when a system is being shilled
Targeting

- Targeting per user per topic
  (e.g., Syskill & Webert)

- Targeting by user
  (e.g., MyYahoo!)

- Targeting by topic
  (e.g., IndexFinder)

- Targeting all users
  (e.g., Top N lists, FAQs, handpicked web sites)
Broadening Issues

- Evaluation
  - Functional vs. human-oriented evaluations
  - Is there something in between?
- Targeting
  - Answers the question ‘for whom are we building this system?’
- Privacy and trust
  - Broader than one user and one system
  - Concept of a weak-tie
- Shilling
  - Involves inundating the system with data intended to coerce it to artificially recommend the perpetrator’s products more often than those of a competitor.
  - Algorithms to detect when a system is being shilled
Conclusions

• Recommenders are systems that connect people.

• The question is ‘do they bring people together by explicitly or implicitly modeling them?’

• Approaches for discovering self-organizing social networks constitute the primary thrust in current RS research.

• Evaluation is challenging with the human in the loop.

• We are trying to make a science out of recommendation.
Recommender Systems Research

Saverio Perugini
Department of Computer Science
University of Dayton

saverio@udayton.edu
http://homepages.udayton.edu/~perugisa