Serendipitous Recommendation: Especially for Scholarly Papers and Mobile Apps

Kazunari Sugiyama

National University of Singapore
Today’s talk

Outline of Singapore

Serendipitous Recommendation for Scholarly Papers

Serendipitous Recommendation for Mobile Apps
Outline of Singapore

- **Population:** 5.60 million (in June 2016)
  - Chinese 74%, Malay 13%, Indian 9%, Others 4%


- **Area**
  - **NUS:** National University of Singapore
  - **NTU:** Nanyang Technological University
  - **SMU:** Singapore Management University
  - **SUTD:** Singapore University of Technology and Design

- **Language**
  - Malay, Mandarin, English, Tamil

Peking University
22.9 times larger than Singapore
Serendipitous Recommendation for Scholarly Papers

- Kazunari Sugiyama and Min-Yen Kan: “Scholarly Paper Recommendation via User's Recent Research Interests” (JCDL’11, short paper)
Introduction

Content-based Recommendation

Candidate papers to recommend

Papers similar to each user’s profile are recommended.

Recommended papers
Introduction

Junior researcher: Only a few recently published paper

Senior researcher: Several past published papers

Serendipitous recommendation is important.

Publication list

('15)    ('16)

Seek to apply their knowledge towards other areas.

Publication list

('01)    ('02)    ('11)

Broaden their range of research interests.
Introduction

When do we find something serendipitous?

Advice from colleagues

That’s nice idea!

How about using this technique?

Attend seminars

We can also use this approach even if the research topic is different from ours!

User profile construction for serendipitous recommendation with others:

- Dissimilar users
- Co-author network

Serendipitous discovery: Interactions with others play an important role.
(1) User profile construction

(1-1) Construction of basic user profile $P_u$

[A researcher’s published papers]

Publication list

<table>
<thead>
<tr>
<th>Old</th>
<th>New</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1$ (‘02)</td>
<td></td>
</tr>
<tr>
<td>$p_2$ (‘03)</td>
<td></td>
</tr>
<tr>
<td>$p_i$ (‘07)</td>
<td></td>
</tr>
<tr>
<td>$p_n$ (‘16)</td>
<td></td>
</tr>
</tbody>
</table>

$W_{p_{n-1}}$ $W_{p_{n-2}}$ $W_{p_{n-1}}$

References

Citation papers

$P_{c_1 \rightarrow p_i}$ (‘07)
$P_{c_2 \rightarrow p_i}$ (‘09)
$P_{c_k \rightarrow p_i}$ (‘15)

$W_{P_{c_1 \rightarrow p_i}}$ $W_{P_{c_2 \rightarrow p_i}}$ $W_{P_{c_k \rightarrow p_i}}$

Reference papers

$P_{i \rightarrow ref_1}$ (‘05)
$P_{i \rightarrow ref_2}$ (‘06)
$P_{i \rightarrow ref_i}$ (‘03)

$W_{P_{i \rightarrow ref_1}}$ $W_{P_{i \rightarrow ref_2}}$ $W_{P_{i \rightarrow ref_i}}$

(1-2) Construction of user profile $P_u^{srdp}$

to recommend serendipitous papers

by using information obtained from:

- **Dissimilar users** (“DU”)
- **Co-author network** (“CAN”)

(2) Feature vector construction for candidate papers $G^{P_{rec}}$

Citation papers

$P_{rec_1}$
$P_{rec_2}$
$P_{rec_3}$

$P_{rec_{-1}}$

$G^{P_{rec_1}}$ $G^{P_{rec_2}}$ $G^{P_{rec_3}}$

Reference papers

$P_{rec_j}$

$W_{P_{rec_j \rightarrow ref_1}}$ $W_{P_{rec_j \rightarrow ref_2}}$ $W_{P_{rec_j \rightarrow ref_1}}$

(3) Recommendation of serendipitous papers

Computing cosine similarity between

$P_u^{srdp}$ and $G^{P_{rec}}$ constructed in (1) and (2),
respectively
Basic User Profile Construction

Weighting scheme
- Cosine similarity

Forgetting factor
User Profile Construction via Dissimilar Users (DU)

User profile generated from published history of papers

User profile for serendipitous recommendation

User 1
User 4 (Sim: 0.16) Weight: 1/(0.16+1)
User 10 (Sim: 0.26) Weight: 1/(0.26+1)

User 2
User 5 (Sim: 0.21) Weight: 1/(0.21+1)
User 1 (Sim: 0.32) Weight: 1/(0.32+1)

User 3
User 1 (Sim: 0.14) Weight: 1/(0.14+1)
User 7 (Sim: 0.25) Weight: 1/(0.25+1)

User n
User 6 (Sim: 0.07) Weight: 1/(0.07+1)
User 2 (Sim: 0.12) Weight: 1/(0.12+1)
User Profile Construction via Co-author Network (CAN)

Consider only radial network from the target researcher, “Y.I. Lin”

### Weighting scheme

1. **Linear Combination (LC)**
2. **Reciprocal of Path Length (RCP-PL)**
3. **Reciprocal of Similarity (RCP-SIM)**
4. **Product of W2 and W3 (RCP-PLSIM)**
### Experiments

#### Experimental Data

(Basic dataset has been released from

(a) **Researchers (they have publication lists in DBLP)**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of researchers</td>
<td>50</td>
</tr>
<tr>
<td>Average number of DBLP papers</td>
<td>10.0</td>
</tr>
<tr>
<td>Average number of relevant</td>
<td>75.4</td>
</tr>
<tr>
<td>papers in our dataset</td>
<td></td>
</tr>
<tr>
<td>Average number of citations</td>
<td>14.8 (max. 169)</td>
</tr>
<tr>
<td>Average number of references</td>
<td>15.0 (max. 58)</td>
</tr>
</tbody>
</table>

(b) **Candidate papers to recommend (constructed from ACM Digital Library)**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of papers</td>
<td>100,351</td>
</tr>
<tr>
<td>Average number of citations</td>
<td>17.9 (max. 175)</td>
</tr>
<tr>
<td>Average number of references</td>
<td>15.5 (max. 53)</td>
</tr>
</tbody>
</table>
Experiments

Evaluation Measure

- **NDCG@5, 10** [Järvelin and Kekäläinen, SIGIR’00]
  - Gives more weight to highly ranked items
  - Incorporates different relevance levels through different gain values
    - 1: Relevant search results
    - 0: Irrelevant search results

- **MRR** [Voorhees, TREC-8, ’99]
  - Provides insight in the ability to return a relevant item at the top of the ranking

- **Normalized Item Novelty@10**
  - Improved “Item Novelty” in [Zhang and Hurley, RecSys’08]
Normalized Item Novelty

Item novelty ($ITN$) [Zhang and Hurley, RecSys’08]

$$ITN = \frac{1}{N} \sum_{j=1}^{N} d(P_{u}^{srdp}, F^{P_{recj}})$$

- $P_{u}^{srdp}$: User profile
- $F^{P_{recj}}$: Feature vector of the candidate paper to recommend

Monotone increasing

Normalized item novelty ($nITN$)

$$nITN = \frac{1}{N} \sum_{j=1}^{N} \frac{d(P_{u}^{srdp}, F^{P_{recj}})}{\max d(P_{u}^{srdp}, F^{P_{recj}})}$$

Avoid monotone increasing
Baselines

- Random Selection
- Maximal Marginal Relevance (MMR)  
  [Carbonell and Goldstein, SIGIR’98]
  - Ranks documents for a given query based on their similarities to the query and also their dissimilarities to other selected documents:

\[
MRR \overset{\text{def}}{=} \arg \max \left[ \lambda \left( Sim_1(D_i, Q) \right) - (1 - \lambda) \max Sim_2(D_i, D_j) \right]
\]

[Carbonell and Goldstein, SIGIR’98]

- Reciprocal of rank \( D_i \) for \( Q \)
- Standard cosine similarity

[Rafiei et al., WWW2010]

MMR-Rafiei(+)

WING, NUS
5-fold cross validation

For 1 user,
- Training: 39 users
- Test: 10 users

User profiles that contain a variety of topics can be constructed by using more dissimilar users.
Results Obtained by Co-author Network (CAN)

[Item novelty@10]

5-fold cross validation

Fine-grained control of weight can be achieved by cosine similarity.
## Comparison of Recommendation Accuracy

<table>
<thead>
<tr>
<th>Method</th>
<th>nDCG@10</th>
<th>MRR</th>
<th>nITN@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random selection</td>
<td>0.127</td>
<td>0.087</td>
<td>0.078</td>
</tr>
<tr>
<td>MMR-Rafiei(+) in (DU) (λ=0.63)</td>
<td>0.372</td>
<td>0.553</td>
<td>0.581</td>
</tr>
<tr>
<td>Our proposed approach (DU) (DU-title, Ndu=8)</td>
<td>0.414</td>
<td>0.612</td>
<td>0.642</td>
</tr>
<tr>
<td>MMR-Rafiei(+) in (CAN) (λ=0.68)</td>
<td>0.353</td>
<td>0.544</td>
<td>0.568</td>
</tr>
<tr>
<td>Our proposed approach (CAN) (RCP-PLSIM, pl=3)</td>
<td>0.426</td>
<td>0.624</td>
<td>0.656</td>
</tr>
</tbody>
</table>
Examples of Serendipitous Recommendation

- Work on user’s search behavior in information retrieval
- Often survey papers in information retrieval-related conferences only

<table>
<thead>
<tr>
<th></th>
<th>nDCG@10</th>
<th>MRR</th>
<th>nITN@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>DU</td>
<td>0.171</td>
<td>0.235</td>
<td>0.513</td>
</tr>
</tbody>
</table>
Examples of Serendipitous Recommendation

- Work on user’s search behavior in information retrieval
- Often survey papers in information retrieval-related conferences only

<table>
<thead>
<tr>
<th></th>
<th>nDCG@10</th>
<th>MRR</th>
<th>nITN@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>DU</td>
<td>0.171</td>
<td>0.235</td>
<td>0.513</td>
</tr>
<tr>
<td>CAN</td>
<td>0.459</td>
<td>1.000</td>
<td>0.618</td>
</tr>
</tbody>
</table>

Our approach can recommend:
- Collaborative research
- Human factors papers, and so on
Examples of Serendipitous Recommendation

• Work on sentiment analysis
• Often survey natural language processing field only

<table>
<thead>
<tr>
<th></th>
<th>nDCG@10</th>
<th>MRR</th>
<th>nITN@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>DU</td>
<td>0.218</td>
<td>0.246</td>
<td>0.497</td>
</tr>
</tbody>
</table>
Examples of Serendipitous Recommendation

- Work on sentiment analysis
- Often survey natural language processing field only

<table>
<thead>
<tr>
<th></th>
<th>nDCG@10</th>
<th>MRR</th>
<th>nITN@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>DU</td>
<td>0.218</td>
<td>0.246</td>
<td>0.497</td>
</tr>
<tr>
<td>CAN</td>
<td>0.463</td>
<td>1.000</td>
<td>0.632</td>
</tr>
</tbody>
</table>

Can recommend
- Human factors
- User interface
- User services papers, and so on
Serendipitous Recommendation for Mobile Apps

- Upasna Bhandari, Kazunari Sugiyama, Anindya Datta, and Rajni Jindal: “Serendipitous Recommendation for Mobile Apps Using Item-Item Similarity Graph” (AIRS 2013)

- Jovian Lin, Kazunari Sugiyama, Min-Yen Kan, and Tat-Seng Chua: “Scrutinizing Mobile App Recommendation: Identifying Important App-related Indicators” (AIRS 2016@Tsinghua Univ., to be presented on 1st Dec.)
Introduction

Statistics on Apple’s iOS App Store
- Provides 550,000 unique apps in 123 countries
- 25 billion downloads

(“Apple’s App Store Downloads Top 25 Billion”:
Introduction

Existing recommendation systems

• Alleviate information overload
e.g.) Amazon.com

Problems

• Users often observe that all of these items are already known.
• Existing recommendation systems often recommend items that the users have rated or downloaded before.
Introduction

• Serendipitous recommendation systems
  • Provide something unexpected

• Our approach to serendipitous recommendation
  • Provide something diverse and novel
  • Achieve this by increasing item novelty and diversity
  • Construct item-item similarity graph using apps installed on each user’s mobile phone
Proposed Method

System overview

App stores → Preprocessing

App stores

Preprocessing

App stores

App ID

Title, comments, reviews (“app summary”)

M1: Similarity Calculation

M2: App-App Similarity Graph Construction

M3: Serendipitous Recommendation Generation

Apps on user’s mobile phone

User

Recommended serendipitous apps
Proposed Method

M1: Similarity calculation

(1) Transform app title and description into TF-IDF feature vector

\[ f^{App_i} = (t_1, t_2, \cdots, t_k) \]

(2) Construct similarity matrix

<table>
<thead>
<tr>
<th></th>
<th>App 1</th>
<th>App 2</th>
<th></th>
<th>App N-1</th>
<th>App N</th>
</tr>
</thead>
<tbody>
<tr>
<td>App 1</td>
<td>--</td>
<td>0.325</td>
<td></td>
<td>0.223</td>
<td>0.512</td>
</tr>
<tr>
<td>App 2</td>
<td>0.325</td>
<td>--</td>
<td></td>
<td>0.452</td>
<td>0.137</td>
</tr>
<tr>
<td>App N-1</td>
<td>0.223</td>
<td>0.452</td>
<td>--</td>
<td>--</td>
<td>0.376</td>
</tr>
<tr>
<td>App N</td>
<td>0.512</td>
<td>0.137</td>
<td>0.376</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>
Proposed Method

M2: App-App similarity graph construction
M3: Serendipitous recommendation generation
Experiments
Experimental Data
• 66,223 apps
• 22,213 users

Collected from
- Apple iTunes
- Google Android
- Blackberry native store
- Windows app store
Experiments

Evaluation measure

- Normalized Item Novelty
- Diversity-in-top-$N$
  - How many distinct apps are recommended in the top-$N$ across all users
Recommendation Accuracy

[Item novelty]

[ Diversity-in-top-N]
Distribution of Item Novelty for Sampled Users

[User 1 (15 apps)]

[User 2 (7 apps)]

[User 1 (40 apps)]
Summary of Today’s Talk

• **Serendipitous Recommendation for Scholarly Papers**
  - Propose constructing user profile for serendipitous recommendation that considers relations among researchers
    - Dissimilar users
    - Co-author network

• **Serendipitous Recommendation for Mobile Apps**
  - Propose serendipitous recommendation for mobile apps by focusing on *item novelty* and *diversity*
  - Construct item-item similarity graph using apps installed on each user’s mobile phone

*Thank you very much!*