Cross Social Media Recommendation

Xiaozhong Liu, Indiana University Bloomington
Tian Xia, Renmin University
Yingying Yu, Dalian Maritime University
Chun Guo, Indiana University Bloomington
Yizhou Sun, Northeastern University

@ICWSM16
If Social Media can Mirror this World...

What the world may look like?
If Social Media can Mirror this World…

What the world may look like?

Social Media Under Fire
Countries that block Twitter, Facebook, or YouTube*

*Till or partial blocks as of March 2014. Sources: Google, Twitter, OpenNet Initiative.
21.97%
As a result, if we are interested to model the world level knowledge, all the research findings based on a single social media system can be **BIASED**, and the social networks or knowledge networks generated from a single system, or to say a single community, **CANNOT** fully represent the people from all of the world.
25% of all social network users in the world are Chinese.
Language Bubbles, Culture Bubbles, and Network Bubbles

Bubbles

Goal: give everyone SAME information access
Enable global information recommendation across different social media - *Pseudo Global Social Media Network (PGSMN)*
Message (content): I don’t think I like #iphone6… nothing new…

Hashtag (vector): #icwsm

User (LM):

1. No physical links between Twitter and Weibo
2. User cannot register both (policy + language barrier)
Message (content): I don’t think I like #iphone6… nothing new…

Explicit Semantic Analysis (ESA)

random walk via PGSMN
Wait!

Twitter + Weibo text can be very short and noisy.

obama…… Barack Obama

Wikipedia redirect page

Explicit Semantic Analysis (ESA)

English text

iPhone

Apple Inc.

iPad

random walk via PGSMN

Chinese text
Explicit Semantic Path Mining (ESPM)

Project text (can be short and noisy) into a number of ranked Wikipedia category paths

Text → ESA vector → Link Analysis (Wiki category tree) → Wiki category paths (ESPM results)

Original Text: Iraq's most influential Shiite cleric suggested on Friday that Prime Minister Nouri al-Maliki needed to recast his approach or step down, adding his powerful voice to growing criticism of the Shiite-dominated government's leader...

<table>
<thead>
<tr>
<th>Rank</th>
<th>5 ESA results</th>
<th>5 ESPM results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22 January 2007 Baghdad bombings</td>
<td>Society-&gt;Religion and society-&gt;Religion and politics</td>
</tr>
<tr>
<td>2</td>
<td>Clericalism</td>
<td>Politics-&gt;Politics by issue-&gt;Petroleum politics</td>
</tr>
<tr>
<td>3</td>
<td>Sadrist Movement</td>
<td>Politics-&gt;Political philosophy-&gt;Political philosophy by politician</td>
</tr>
<tr>
<td>4</td>
<td>Sipah-e-Muhammad Pakistan</td>
<td>People-&gt;Aspects of individual lives-&gt;Deaths by person</td>
</tr>
<tr>
<td>5</td>
<td>23 November 2006 Sadr City bombings</td>
<td>Politics-&gt;Politics by country-&gt;Politics of Iraq/Iraqi nationalism</td>
</tr>
</tbody>
</table>

Original Text (SportsCenter@SportsCenter Aug 17): Greg Oden, the No. 1 overall pick in 2007 NBA draft, has signed a one-year, $1.2 million deal to play in China.

<table>
<thead>
<tr>
<th>Rank</th>
<th>5 ESA results</th>
<th>5 ESPM results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Greg Oden</td>
<td>Sports-&gt;Sports trophies and awards-&gt;Basketball trophies and awards-&gt;Gatorade National Basketball Player of the Year</td>
</tr>
<tr>
<td>2</td>
<td>Henry Oden</td>
<td>Sports-&gt;Works about sports-&gt;Sports video games-&gt;NBC Sports video games</td>
</tr>
<tr>
<td>3</td>
<td>Oden</td>
<td>Sports-&gt;Sports business-&gt;Sports management companies-&gt;Maple Leaf Sports &amp; Entertainment</td>
</tr>
<tr>
<td>4</td>
<td>William B. Oden</td>
<td>Sports-&gt;Sports-related lists-&gt;Top sports lists-&gt;National Basketball Association statistical leaders</td>
</tr>
<tr>
<td>5</td>
<td>Lon Oden</td>
<td>Sports-&gt;Sports controversies-&gt;Basketball controversies-&gt;National Basketball Association controversies</td>
</tr>
</tbody>
</table>

Input can be a #hashtag (associated text)

#iPadAir2

1. Technology->Mobile operating systems->IOS (Apple) (P=0.45)
2. Technology->Personal computers->Tablet computers (P=0.55).
Explicit Semantic Path Mining (ESPM)

26 general categories, i.e., Culture, Education, Environment, Politics, and Science

Wikipedia Category Tree
(871,978 nodes + 1,229,833 edges)

Input: text
Output: Wikipedia category path

American military personnel killed in the War of 1812
Tree like category graph

\[ P(C_i|t) = \sum_{C_i \rightarrow C_{child_k}} \frac{P(C_{child_k}|t)}{|C_j \rightarrow C_{child_k}, \forall j| \cdot w^{\text{norm}}(C_i|t)} \]

Input: text

0.755 \rightarrow 0.51 (0.65, 0.51)

0.49 (0.18, 0.14)

0.35 (0.45, 0.35)

0.245

\[ w(c_j|t) = \frac{\sum_{a_k \in c_j} \left( \lambda \cdot p(a_i|t) + (1 - \lambda) \cdot \frac{p(a_k|t)}{|a_i \in c_j|} \right)}{|a_k \in c_j|} \]
Random Walk between different category paths (on the tree)

when $t > 1$

$$RW(P_{i(t)} \sim P'_{j(t)}) = \alpha \cdot RW(C_{i(t)} \sim P'_{j(t)}) + (1 - \alpha) \cdot P(C_{i(t+1)}|C_{i(t)}) \cdot RW(P_{i(t+1)} \sim P'_{j(t+1)})$$

when $t = 1$

$$RW(P_i \sim P'_j) = RW(P_{i(1)} \sim P'_{j(1)})$$

Decay function for random walk
- higher level (expert contributed) can be more important
Random Walk between Twitter and Weibo (case: recommend twitter hashtag to Weibo users)

<table>
<thead>
<tr>
<th>ID</th>
<th>Meta-path</th>
<th>Ranking Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>$U_{Weibo} \rightarrow HT_{Weibo} \rightarrow A \rightarrow HT_{Twitter}$</td>
<td>The candidate Twitter HashTag is important, if it is relevant to the same Wikipedia page as the query Weibo User.</td>
</tr>
<tr>
<td>F2</td>
<td>$U_{Weibo} \rightarrow HT_{Weibo} \rightarrow A \rightarrow HT_{Twitter}$</td>
<td>The candidate Twitter HashTag is important, if it is mentioned by the Twitter User who is relevant to the same Wikipedia page as the query Weibo User.</td>
</tr>
<tr>
<td>F3</td>
<td>$A \rightarrow HT_{Twitter}$</td>
<td>The candidate Twitter HashTag is important, if its co-occur HashTag is relevant to some Wikipedia page which links to the one relevant to the Weibo HashTag which is mentioned by the query Weibo User.</td>
</tr>
<tr>
<td>F4</td>
<td>$U_{Weibo} \rightarrow HT_{Weibo} \rightarrow A \rightarrow HT_{Twitter}$</td>
<td>The candidate Twitter HashTag is important, if it is mentioned by the Twitter User who is relevant to the same Wikipedia page which links to the one relevant to the Weibo HashTag which is mentioned by the query Weibo User.</td>
</tr>
<tr>
<td>F5</td>
<td>$U_{Weibo} \rightarrow HT_{Weibo} \rightarrow A \rightarrow HT_{Twitter}$</td>
<td>The candidate Twitter HashTag is important, if its relevant Wikipedia page’s category is the same as the query Weibo User’s mentioned hashtag.</td>
</tr>
<tr>
<td>F6</td>
<td>$A \rightarrow HT_{Twitter}$</td>
<td>The candidate Twitter HashTag is important, if its relevant Wikipedia page’s category has category which has page that is relevant to the query Weibo User.</td>
</tr>
<tr>
<td>F7</td>
<td>$A \rightarrow HT_{Weibo} \rightarrow A \rightarrow HT_{Twitter}$</td>
<td>The candidate Twitter HashTag is important, if its co-occur HashTag is relevant to some Wikipedia page which belongs to the same category as the one relevant to the Weibo HashTag which is mentioned by the query Weibo User.</td>
</tr>
<tr>
<td>F8</td>
<td>$A \rightarrow HT_{Weibo} \rightarrow A \rightarrow HT_{Twitter}$</td>
<td>The candidate Twitter HashTag is important, if its relevant category path is related to the query Weibo User’s relevant category path.</td>
</tr>
<tr>
<td>F9</td>
<td>$A \rightarrow HT_{Weibo} \rightarrow A \rightarrow HT_{Twitter}$</td>
<td>The candidate Twitter HashTag is important, if its co-occur HashTag’s relevant category path is related to the query Weibo User’s mentioned hashtag’s relevant category path</td>
</tr>
<tr>
<td>F10</td>
<td>$A \rightarrow HT_{Weibo} \rightarrow A \rightarrow HT_{Twitter}$</td>
<td>The candidate Twitter HashTag is important, if it is mentioned by some Twitter User whose relevant category path is related to the query Weibo User’s related users’ relevant category path</td>
</tr>
<tr>
<td>F11</td>
<td>$A \rightarrow HT_{Weibo} \rightarrow A \rightarrow HT_{Twitter}$</td>
<td>The candidate Twitter HashTag is important, if its relevant category path is related to the query Weibo User’s mentioned hashtag’s relevant category path</td>
</tr>
</tbody>
</table>

Learning to Rank
Experiment

Raw Data:
2012/09/17 - 2012/09/23 (1 week data).

Weibo: 3,296,945 messages
Twitter: 20,128,826 messages

Wikipedia March 2014 Dump

Testing set:
459 hashtags shared by Twitter and Weibo users

401 Weibo users use those Hashtag

Remove those from Weibo for evaluation (20,248 user-hashtag pairs for evaluation)
Experiment - baseline

Google Machine Translation:

Weibo user -> text -> translate to English -> BM25 ->
Twitter hashtag index  -> top ranked Twitter hashtags

Language can be noisy… spoken language… challenge machine learning algorithms…
Result

MRR

Machine Translation
Wiki-page
Wiki-category
Wiki-page+category
ESPM
ALL
Categories: The Voice (TV series)
Conclusion

Cross-social media information recommendation is a novel but important task

Facebook <-> Twitter (user has accounts for both)
Twitter <-> Weibo (user cannot register for both)

Wikipedia can be a bridge

Random walk via Pseudo Global Social Media Network (PGSMN)
Future

Community-based Information Adoption/Diffusion Comparison of Different Microblogging Sites – ACM Hypertext 2016

Twitter

Weibo

Source Data

Social network of authors.
- Show Twitter and Weibo
- Show only Twitter
- Show only Weibo

Nodes
- Show Users
- Show Keywords / Hashtags

Keywords / Hashtags
- Hashtags
- Keywords

Statistics
Thank you.

liu237@indiana.edu