Recommending Information Sources to Information Seekers in Twitter

Marcelo Armentano, Daniela Godoy and Analía Amandi

ISISTAN Research Institute, UNCPBA
CONICET, Consejo Nacional de Investigaciones Científicas y Técnicas
Tandil, Argentina
Twitter

• Non-directed connections
  – 77.9% of connections are unidirectional [1]
• 140-characters-length posts
• Tree categories of users [2,3]
  • Information Sources
  • Information Seekers
  • Friends

• Traditional recommender systems
  – ratings

• Information available for profiling users:
  – follow/followee relationships
  – tweets
Followees Recommendations on Twitter

• Identify users posting relevant tweets for a target user

• User Profiles
  – Strategy T0
  – Strategy T1
  – Strategy T2

• Topology based search of candidates
Profile Strategy T0

“Users tweet about things that interest them”

tweets(u_T) = \{t_1, t_2, \cdots, t_k\}

t_j = \{w_{1,j}, w_{2,j}, \cdots, w_{t,j}\}

profile_{T0}(u_T) = \{< w_1, tf_{w_1} >, < w_2, tf_{w_2} >, \cdots, < w_n, tf_{w_n} >\}

Filters applied:
1. Remove punctuation symbols (emoticons)
2. NoSlang online dictionary
3. Remove Stopwords
4. Porter stemming algorithm
Profile Strategy T1

“Users follow people that tweet about things that interest them”

\[
\text{followees}(u_T) = \{f_1, f_2, \cdots, f_m\}
\]

\[
\text{profile}_{T1}(u_T) = \{\text{profile}_{T0}(f_1), \text{profile}_{T0}(f_2), \cdots, \text{profile}_{T0}(f_m)\}
\]
Profile Strategy T2

“Users are likely to follow people in different interest categories”

\[ \text{profile}_{T2}(u_T) = \{ fc_1, fc_2, \cdots, fc_p \} \]

\[ c_{fc_i} = \frac{1}{|fc_i|} \sum_{f \in fc_i} \text{profile}_{T0}(f) \]

\[ fc_{f_{new}} = \arg \max_{j=1,...,p} \text{sim}(\text{profile}_{T0}(f_{new}), f_{fc_j}) \]
Topology-based search of candidates

\[ T = \bigcup_{l \in L} \text{followees}(l) \]

\[ L = \bigcup_{s \in S} \text{followers}(s) \]

\[ S = \bigcup_{s \in \text{followees}(U)} S \]

\[ R = T - S \]
Ranking candidates

\[ \text{sim}_{T0}(u_C, u_T) = \text{sim}_{\cosine}(\text{Profile}_{T0}(u_C), \text{Profile}_{T0}(u_T)) \]

\[ \text{sim}_{T1}(u_C, u_T) = \max_{f_i \in \text{followees}(u_T)} \text{sim}_{T0}(f_i, u_C) \]

\[ \text{sim}_{T2}(u_C, u_T) = \max_{c_{f_c_i} \in \text{followees}(u_T)} \text{sim}_{T0}(c_{f_{c_i}}, u_C) \]
Experimental evaluation

• Twitter Dataset
  – Originally available at
  – 835,541 follower/followee relationships
  – 456,107 users
  – 10,467,110 tweets (2006-2009)
  – 500 seeds

Experimental evaluation

- $|U_{\text{test}}|=100$
- $|\text{followees}| > |\text{followers}|$
- Users with less than 10 tweets were not considered

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>Maximum</th>
<th>Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>#followees</td>
<td>94.77±15.54</td>
<td>119</td>
<td>41</td>
</tr>
<tr>
<td>#followers</td>
<td>2.0±1.74</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>#tweets</td>
<td>102.44±57.56</td>
<td>199</td>
<td>11</td>
</tr>
</tbody>
</table>
Experimental evaluation

• Hold-out cross validation
  – 70% training, 30% testing
  – 5 tests for each target user
precision(RE_u) = \frac{1}{|U_{test}|} \sum_{u \in U_{test}} \frac{|followees_{test}(u) \cap RE_u|}{|RE_u|}
Results

Average reciprocal hit rank

\[
\text{ARHR}(\text{RE}_u) = \frac{1}{|U_{test}|} \sum_{i=1}^{h} \frac{1}{p_i}
\]

Position of hits in RE: \( p_1, p_2, \ldots, p_h \)
Results

Average similarity of user profiles

![Bar graph showing average similarity of user profiles](image)
Conclusions

• The user’s own tweets are not a good source of profiling knowledge

• Modelling a user’s interests using the tweets of his/her followees reach high levels of precision in the recommendations
  – Individual vectors for each followee
  – Grouping followees
Current and Future work

• Use more content information
  – linked web pages
  – Hashtags

• Measure the impact of other factors
  – Number of occurrences of a candidate in the final list
  – Identify and filter non information sources
Thanks for your attention!

Marcelo Armentano
marcelo.armentano@isistan.unicen.edu.ar
marcelo.armentano@gmail.com
References
