Generating ordered list of Recommended Items: A Hybrid Recommender System of Microblog

Yingzhen Li and Ye Zhang

School of Mathematics & Computational Science, Sun Yat-sen University, Guangzhou, China

中山大學 SUN YAT-SEN UNIVERSITY

Introduction

- A solution of KDD Cup 2012, track 1 task, which requires predicting users a user might follow in Tencent Microblog.
- Tencent Microblog has some special properties which we'll introduce.
- The system consists of:
  - ▷ keyword analysis
  - ▷ user taxonomy
  - ▷ item ranking
  - ▷ (potential)interests extraction
  - ▷ item recommendation(grading process)

Speciality of Tencent Microblog

- It attracted a lot of registered users and became one of the dominant microblog platforms in China based on the large user group of its instant messaging service QQ.

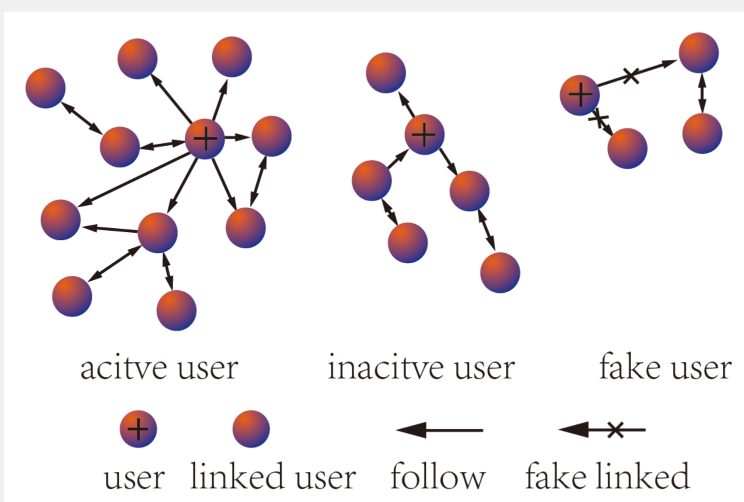
- It is embedded in Tencent's other leading platforms.
- It considers those who frequently write(including retweet and comment) or read microblog messages - no matter on the website or other associated platforms - as active users.

Keyword Analysis

- Applying association rule algorithm to find them directly in the huge keyword set is unrealistic.
- We parallel this process by adopting revised FDM and insert the ambiguous keywords into different classes simultaneously.
- The necessity and sufficient condition for a frequent itemset is the frequency of all its subsets.
- User set: $\mathbf{U} = \{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_m}\}$
- $\mathbf{u_j}$'s keyword set: $\mathbf{K_j} = \{\mathbf{k_{j1}}, \mathbf{k_{j2}}, \ldots, \mathbf{k_{jn_j}}\}$ with weights $\mathcal{W}_j = \{\mathbf{w_{j1}}, \mathbf{w_{j2}}, \ldots, \mathbf{w_{jn_j}}\}$
- $\mathbf{keyword\_class} = \{\mathbf{class_1}, \mathbf{class_2}, \ldots, \mathbf{class_N}\}$, $\mathbf{class_i} = \{\mathbf{k_{i1}}, \mathbf{k_{i2}}, \ldots, \mathbf{k_{im}}\}$.

User Taxonomy

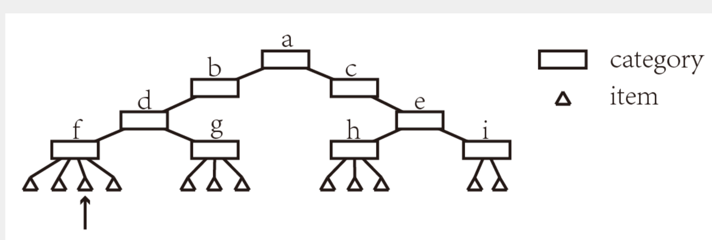
- We divide the users into 3 groups by $\mathbf{user\_class(u_j)}$:
  - ▷ Active - lots of tweets/interactions;
  - ▷ Inactive - few messages/interactions;
  - ▷ **Fake** - they don't login the platform directly, and their messages are synchronized from related platforms.



- $$\mathbf{user\_class(u_j)} = \begin{cases} \text{active}, & \mathbf{act(u_j)} \geq \mathbf{min\_activeness} \\ \text{inactive}, & \mathbf{0} \leq \mathbf{act(u_j)} < \mathbf{min\_activeness} \\ \text{fake}, & \mathbf{act(u_i)} = \mathbf{0} \end{cases}$$
- $\mathbf{act(u_j)} = \mathbf{tweet} \times \mathbf{is\_fake(u_j)}$
- $\mathbf{is\_fake(u_j)} = \dfrac{\mathbf{1} + \mathbf{sgn(at + retweet + comment - min\_action)}}{\mathbf{2}}$

Item Ranking

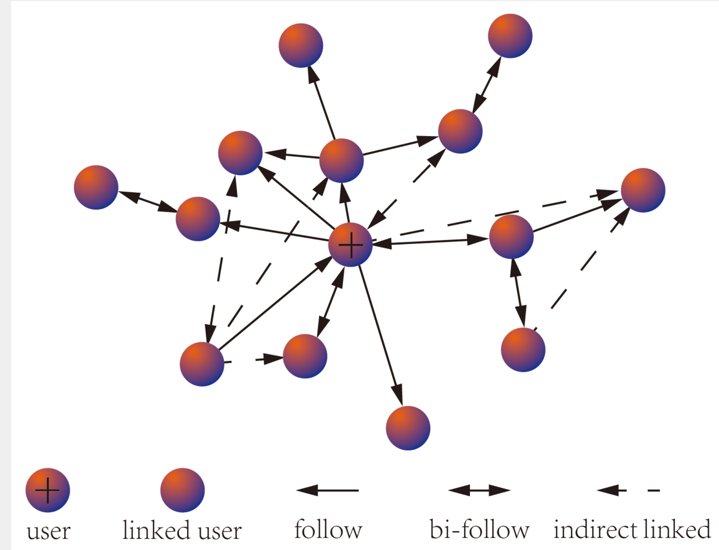
- Items are organized in different categories of professional domains by Tencent to form a hierarchy.
  - ▷ The pointed item belongs to the category a.b.d.f.



- Item set: $\mathbf{I} = \{\mathbf{i_1}, \mathbf{i_2}, \ldots, \mathbf{i_n}\}$ $(\mathbf{i_k} \in \mathbf{h_k})$
- Category set: $\mathbf{H} = \{\mathbf{h_1}, \mathbf{h_2}, \ldots, \mathbf{h_n}\}$
- Counts the number of $\mathbf{i_k}$'s followers and return its ranking in $\mathbf{h_k(I)}$:
  - ▷ The rank of $\mathbf{i_k}$ in $\mathbf{h_k}$: $\mathbf{hot_k} = \mathbf{get\_hot\_rank(i_k, h_k)}$
  - ▷ The rank of $\mathbf{i_k}$ in $\mathbf{I}$: $\mathbf{HOT_k} = \mathbf{GET\_HOT\_RANK(i_k)}$

(Potential)Interest Extraction

- $\mathbf{key\_class(u_j)} = \{\mathbf{class_{ji}}\}$
  - ▷ $\mathbf{W_{ji}} = \sum_{\mathbf{k_l} \in \mathbf{K_j} \cap \mathbf{class_{ji}}} \mathbf{w_l}$
- $\mathbf{potential\_key(u_j)} = \{\mathbf{class_{ji}}\} = \bigcup_{\mathbf{u_k} \in \mathbf{related\_users(u_j)}} \mathbf{key\_class(u_k)}$
  - ▷ $\widetilde{\mathbf{W}}_{\mathbf{ji}} = \sum_{\substack{\mathbf{u_k} \in \mathbf{related\_users(u_j)}, \\ \mathbf{class_{kl_k}} \mathbf{class_{ji}}}} \mathbf{W_{kl_k} fami(u_j, u_k)}$
  - ▷ $\mathbf{fami(u_j, u_k)} = \omega_1 \mathbf{f(at)} + \omega_2 \mathbf{f(retweet)} + \omega_3 \mathbf{f(comment)}$

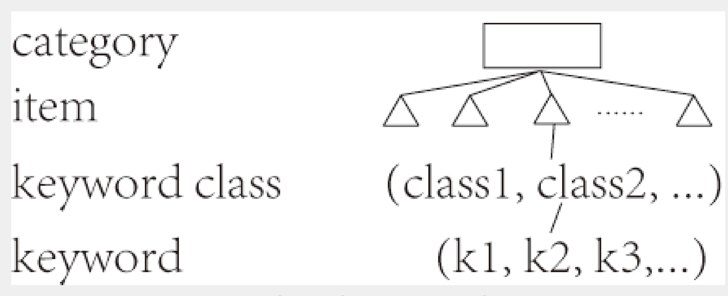


- $\mathbf{interests(u_j)} = \mathbf{key\_class(u_j)} \cap \mathbf{potential\_key(u_j)} = \{\mathbf{class_{jl}}\}$
  - ▷ $$\mathbf{W_{jl}} = \begin{cases} \mathbf{W_{jl}}, & \mathbf{class_{jl}} \in \mathbf{key\_class(u_j)} \\ \widetilde{\mathbf{W}}_{\mathbf{jl}}, & \mathbf{class_{jl}} \in \mathbf{potential\_key(u_j)} \\ \frac{1}{2}(\mathbf{W_{jl}} + \widetilde{\mathbf{W}}_{\mathbf{jl}}), & \mathbf{class_{jl}} \text{ in both sets} \end{cases}$$

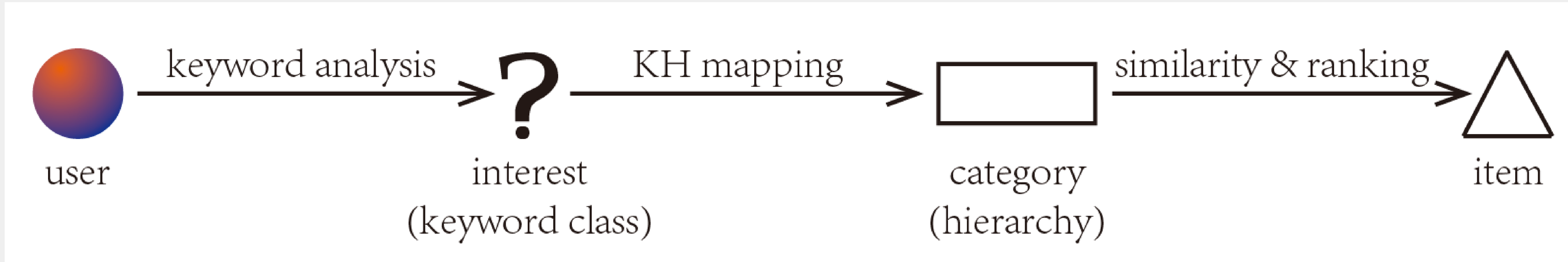
Item Recommendation(Grading Process)

- $\mathbf{KH(h_k)} = \{\mathbf{class_{kj}}\}$
  - ▷ $\hat{\mathbf{W}}_{\mathbf{kp}} = \mathbf{average(W_{jl_j})}$

    $\mathbf{class_{jl_j}} = \mathbf{class_{kp}} \in \mathbf{key\_class(i_j)}$



- $\mathbf{fond(u_j, h_k)} = \mathbf{g(class\_weight(u_j) \cdot class\_weight(h_k), 100)}$
- $\mathbf{grade(u_j, i_k)} = \mathbf{2fond(u_j, h_k)}(\alpha_1 \mathbf{hot_k} + \alpha_2 \mathbf{sim(u_j, i_k)}) - \mathbf{1}$
  - ▷ $\mathbf{sim(u_j, i_k)} = \mathbf{n(|class\_weight(u_j) - class\_weight(i_k)|)}$
  - ▷ $\alpha_1 + \alpha_2 = \mathbf{1}, \alpha_i \geq \mathbf{0}$(obtained in the training process)
- $\mathbf{n(x)}$ and $\mathbf{g(x, y)}$ are the normalization functions.
- The top-3 items are picked out for recommendation.



Results: Training Results of the Parameters

- $\omega_i = \frac{1}{3}$
- α_1 reflects the inclination of accepting popular items.

class	user	followee	interaction	keyword	α_1
active	3919	46	87	10	0.33
inactive	1194	27	42	8	0.18
fake	825	18	2	5	/

Result: Evaluation of the Performance

- Evaluation metric: average precision (which KDD Cup's organizers adopted):

  $$\mathbf{AP@3(u_j)} = \sum_{\mathbf{i=1}}^{\mathbf{3}} \mathbf{p(i)\Delta r(i)}$$

  - ▷ $\mathbf{p(i)}$ is the precision of the $\mathbf{i^{th}}$ recommended item,
  - ▷ $\mathbf{\Delta r(i)}$ is the change in the recall from $\mathbf{i - 1}$ to $\mathbf{i}$.

	active	inactive	fake	total
MAP@3	0.41066	0.46879	0.33606	0.41198

$\mathbf{u_j}$	user class	item	accepted	**AP@3**
2071402	active	1606902 1760350 1774452	1606902 1774452	0.83
942226	inactive	1606902 1606609 1774452	1606902	1.00
193889	fake	1760642 1774684 1774862	1774862	0.33

About

- Yingzhen Li
  Senior Year Undergraduate
  Department of Mathematics
  liyzhen2@mail2.sysu.edu.cn
  http://www.yingzhenli.net/
- Ye Zhang
  Senior Year Undergraduate
  Department of Mathematics
  zhangye5@mail.sysu.edu.cn
  http://dantepy.yslsg.org/