Knowledge Sharing via **Social Login**: Exploiting Microblogging Service for Warming up Social Question Answering Websites

Yang Xiao, Wayne Xin Zhao, Kun Wang, Zhen Xiao
NC&IS LAB, Peking University, China

Outline

- **Motivation**
  - Social Login
  - Cold start in CQA
- Our Method
  - System Design
  - Features
- Experiment
  - Experiment Setup
  - Results Analysis
- Conclusion

The Social World

555 million Users
58 million Tweets
Per Day

1,310,000,000
Active Users
18 minutes Spent
Per Visit

560 million Users

829 million Users
Mobile QQ cover
all smartphones.

220 million Users

Social Login

Third Party Websites

- Users can surf the Internet using **Consistent Identities**
- Collect crowds of users in **Short** time
- Gain **Reliable** user profiles
- Weibo Open API:
  - **600,000** third party websites, **60,000,000** external websites visits daily

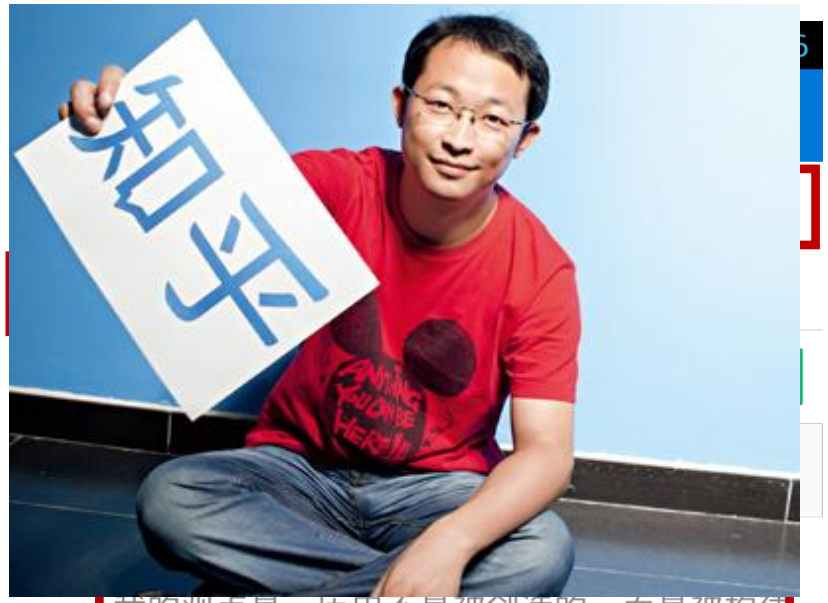
However,

- Despite that social media data is **Abundant**, only **Simple** profiles are leveraged via social login…
- Mine **more value** from social login?

Community Question Answering

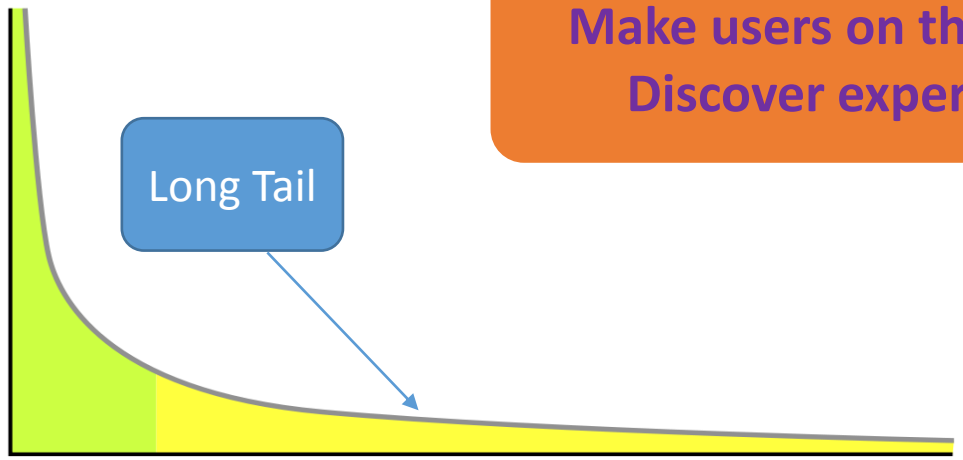
Zhihu Q&A site

- First social network based Q&A
  - User graph
  - User topic graph
  - User question graph
- High-quality questions and answers
  - True domain experts participation
  - Primary experience



Long Tail Phenomenon

- Most contributions in CQA services are made by **a small number of** users.
  - 85% of users answer fewer than 10 questions
  - 60% of users answer fewer than 4 questions
  - Hard to estimate users' expertise
- New comers are prone to leave CQA services very soon



Make users on the long tail stay longer
Discover experts at an early stage

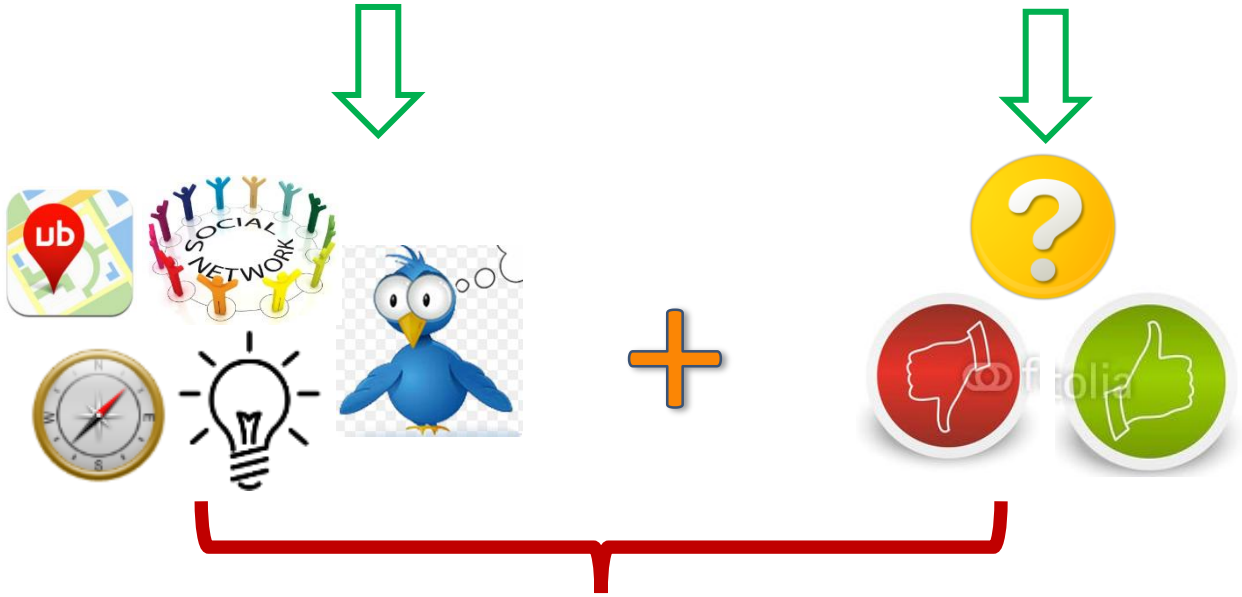
Outline

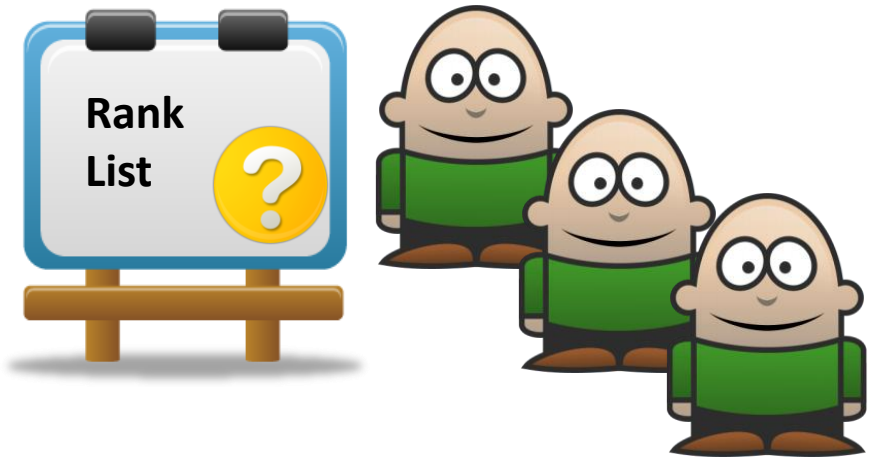
- Motivation
  - Social Login
  - Cold start in CQA
- **Our Method**
  - System Design
  - Features
- Experiment
  - Experiment Setup
  - Results Analysis
- Conclusion

Bridge the gap

Bridge the gap

Candidates Rank = <Weibo Footprints> + <Zhihu Performance>





Weibo and Zhihu Features

Weibo: Relationship Perspective

- Users who have **higher prestige** tends to provide **better** answer
- PageRank
  - $s^{n+1} = \mu M^T s^n + (1 - \mu) y$
- **Performance Biased** Random Walk
  - y represents the user performance on Zhihu

Weibo : Content Perspective

- Users who are **more interested** in the question related topic tends to provide **better** answer
- Model the **relevance** between a question and a user
- Relevance: KL divergence
- $Rel(q,u) = -KL(\theta^q, \theta^u)$

$$= \sum_{\omega \in \mathcal{V}} p(\omega|\theta^q) log \frac{p(\omega|\theta^{q)}}{p(\omega|\theta^u)}$$

Weibo: Content Perspective

- θ^q estimation
  - Question sparsity problem
  - Translation model
  - $\theta_\omega^q \propto \sum_{t \in q} p(\omega, t) = \sum_{t \in q} p(\omega|t) p(t|q)$

- θ^u estimation
  - Tweets accumulation
  - $\theta_\omega^q = \frac{\#(\omega,u)+1}{\sum_{\omega' \in \mathcal{V}} \#(\omega',u)+|\mathcal{V}|}$

Use tags to index questions

Zhihu Features

- In order to test Weibo effect, we take Zhihu features as **baseline**.

Features
Number of Best Answers
Number of Answers
Number of Received Votes
Average Number of Votes
Smoothed Average number of Votes
Best Answer Ratio
Smoothed Best Answer Ratio
Average Answer Length

Outline

- Motivation
  - Social Login
  - Cold start in CQA
- Our Method
  - System Design
  - Features
- **Experiment**
  - Experiment Setup
  - Results Analysis
- Conclusion

Experiment Setup

Dataset

- Crawling Zhihu
  - Snowball-crawled Webpages
  - 266K users, 819K questions, 2.7 million answers
  - **50%** of users log in using **Sina Weibo** account
- Crawling Weibo
  - Crawl the linked users' weibo pages and relationships
- Dataset

Users	Questions	Answers	Topics	Tweets
20,742	335,145	883,373	44,333	21,121,955

Hypothesis testing

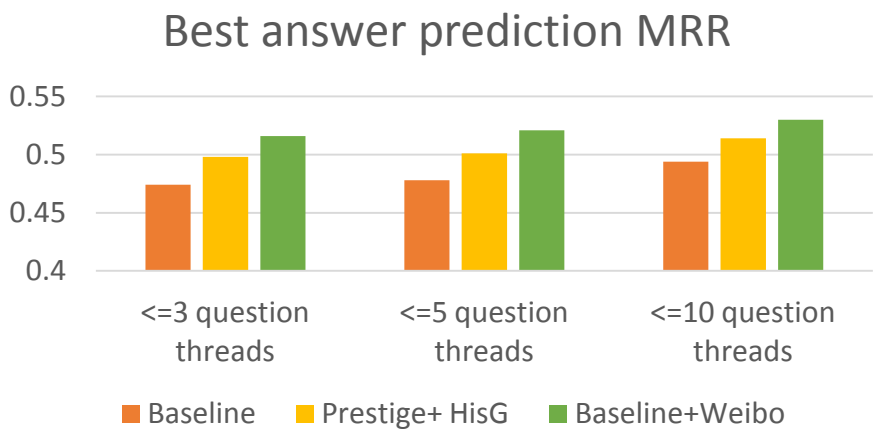
- Spearman Correlation Test
- Prestige
  - Grouping users into buckets
  - Rho = 0.561
- Relevance
  - 14.48% question threads conveys that relevance is correlated with user performance.

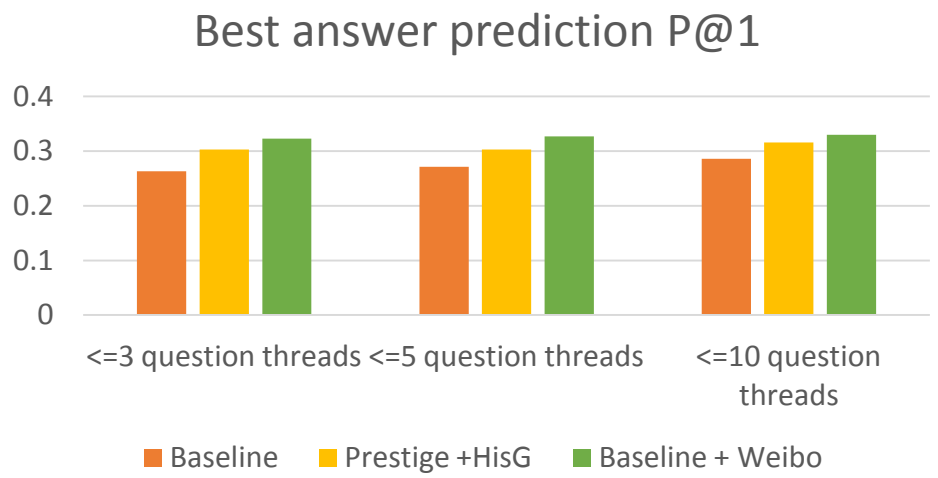
Experiment Setup

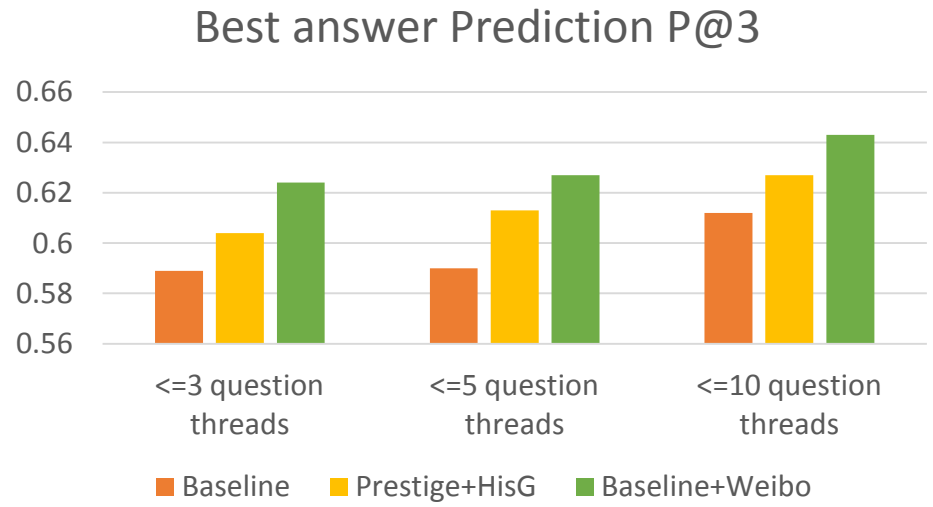
- Tasks:
  - Best Answer prediction
  - User ranking prediction
- Ground Truth
- SVMRank is adopted as the learning framework
- Evaluation Metrics:
  - P@n
  - MRR
  - NDCG@n

Experiment Results

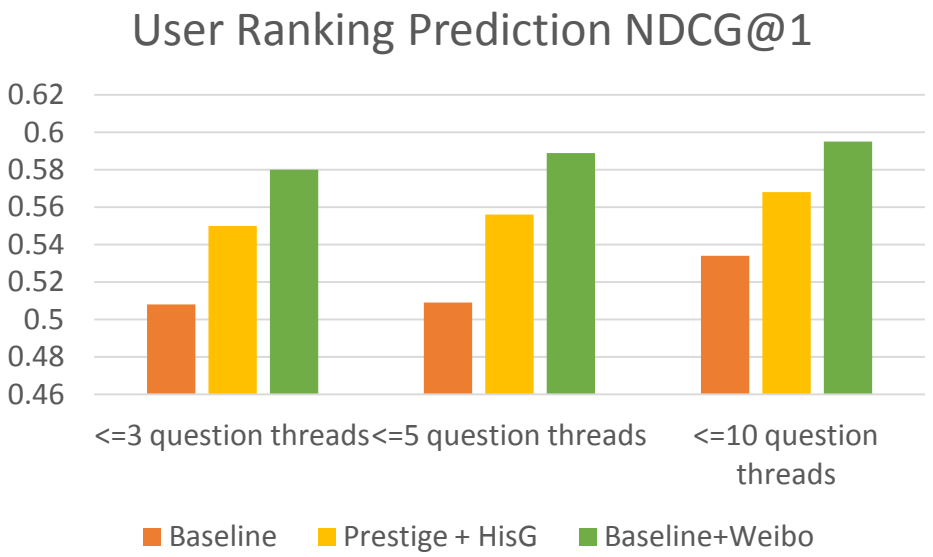
Experiment Results

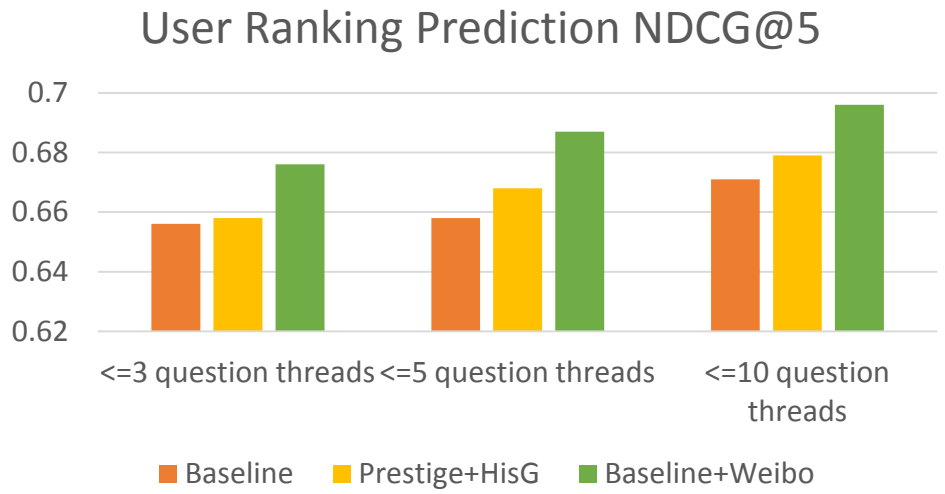


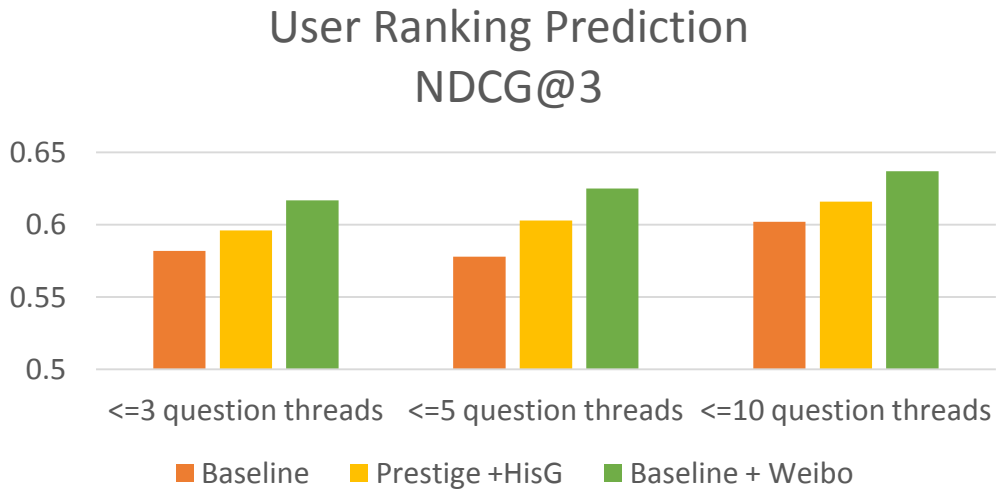




Experiment Results







Outline

- Motivation
  - Social Login
  - Cold start in CQA
- Our Method
  - System Design
  - Features
- Experiment
  - Experiment Setup
  - Results Analysis
- **Conclusion**

Conclusion

- Weibo knowledge is effective to improve prediction results on Zhihu
- Scalability
  - Recommendation system also experience serious cold start problem
  - The method can extend to many other third party startup websites to boost the system

Q&A