iASK: A Distributed Q&A System
Incorporating Social Community and Global Collective Intelligence
Guoxin Liu and Haiying Shen

Presenter: Haiying Shen
Associate professor

*Department of Electrical and Computer Engineering, Clemson University, Clemson, USA
Outline

- Introduction
- Related work
- iASK design
- iASK implementation
- Evaluation
- Conclusion
Introduction

• Vital role of Web Q&A
  ◦ Yahoo! Answers
    • 10 million users in first 2 years
    • Currently 200 million users
    • 15 million visits everyday
  ◦ Drawbacks
    • Unsolved non-factual questions without knowing personnel preferences
    • Long delay due to too many questions needed to be browsed
    • Lack of trustworthiness
Introduction

- Social-based Q&A
  - Potential benefits
    - Personnel recommendation/opinion
    - Trustable and altruistic
  - Problem
    - Confine the Q&A activities within individual social communities
  - Challenge
    - How to connect different social communities for users to efficiently receive answers outside of their social communities
Introduction

• Our Approach:
  ◦ iASK: a unified system that incorporates social community intelligence and global collective intelligence into a single distributed Q&A system
    • A neural network based friend ranking method to identify answerer candidates in the social network
    • A virtual server tree in the central servers to efficiently locate answerer candidates in the global user base
    • A fine-grained reputation system to accurately locate cooperative global experts to answer questions
Outline

- Introduction
- Related work
- iASK design
- iASK implementation
- Evaluation
- Conclusion
Related work

- Social-based Q&A
  - Infrastructure
    - Centralized solutions
    - High overhead for computing
    - Distributed Q&A system
    - Flooding: high communication overhead
    - Selecting: lack of cooperation of global collective intelligence
  - Expert locating algorithm
    - Social features
    - Answerer reputation
    - Question quality
Outline

• Introduction
• Related work
• iASK design
• iASK implementation
• Evaluation
• Conclusion
iASK Design

• Design rationale and challenge
  ◦ Questions inside social community
    • Social intelligence
      • Share similar interests
      • Know friends’ background
      • Need to be accurate and efficient
  ◦ Questions outside social community
    • Global collective intelligence
      • Need to ensure timely and high-quality answers
iASK Design

- iASK architecture
  - Clustering: interest-based virtual server tree
  - Social intelligence: bi-direction friendship
  - Global intelligence: follower-followee

Social community intelligence

iASK's social communities

- $V_P$: Pop
- $V_F$: Folk music
- $V_R$: R.A.P.
- $V_C$: Classical
- $V_N$: News

Global Collective intelligence

- Root
- Music
- Television
- Classical
- Folk music
- Pop
- R.A.P.
iASK Design

- Social intelligence: inside asker’s social communities
  - Neural network-based friend ranking
    - Hidden layer
      - Efficiency: cooperativeness
      - Accuracy: answer quality
    - First layer
      - Response rate/delay + mutual interaction frequency + precision rate

\[ w_1 w_2 \ldots w_{10} \]

\( W \): influence weight

Answer QoS

Cooperativeness

Answer quality

Response rate

Mutual interaction frequency

Response delay

Precision rate
iASK Design

- Global intelligence: outside asker’s social communities
  - Efficiency: interest-based clustering for all users
  - User join/leave: have a new interest/remove an old interest
  - Virtual server: global intelligence collection

![Diagram of iASK Design]

- $V_{1,1}$: Music
- $V_{1,5}$: Research
- $V_{1,n}$: Sports
- $V_{2,1}$: Pop music
- $V_{2,40}$: Datacenter
- $V_{i,j}$: user (sub)$^{i-1}$-interest $j$
- $V_{i,m}$: user (sub)$^{i-1}$-interest $m$
iASK Design

- Fine-grained reputation-based answerer selection
  - Ranking: global reputation + specific expertise
    \[
    R_{u_j} = \frac{1}{\frac{1}{2} \left( \frac{1}{R^g_{u_j}} + \frac{1}{R^{I_i}_{u_j}} \right)}
    \]
  - Global reputation: expertise + followees’ reputation
    \[
    R^g_{u_j} = \frac{1}{\frac{1}{2} \left( \frac{1}{B_{u_j}} + \sum_{u_i \in f(u_j)} \frac{1}{R^g_{u_i} / |f(u_j)|} \right)}
    \]
  - Specific expertise
    \[
    R^{I_i}_{u_j} = \frac{N^{I_i}_{u_j}}{N^{I_i}}
    \]
Outline

- Introduction
- Related work
- iASK design
- iASK implementation
- Evaluation
- Conclusion
iASK implementation

- Two different roles:
  - Virtual server side
    - Java servlet + Tomcat 7.0 + MySQL
  - User side
    - Java applet framework

- Functionality: menu + ask + answer
Outline

- Introduction
- Related work
- iASK design
- iASK implementation
- Evaluation
- Conclusion
Evaluation

- **Experimental settings**
  - 100,000 users
    - Question and answer activity from Yahoo! Answer [1]
    - Social relationship from Facebook trace [2]
  - 100 questions per user

- **Measured metric**
  - Response rate
  - Recall rate: $|\text{RA} \cap \text{BA}| / |\text{BA}|$
  - Precision rate: $|\text{RA} \cap \text{BA}| / |\text{RA}|$
  - Response delay

Evaluation

• Comparison methods
  ◦ Social intelligence
    • Random: randomly select friend
    • Flooding: select all friends
    • SOS [1]: social closeness plus interest similarity
  ◦ Social plus global intelligence
    • Global(Tree): use global intelligence only
    • Global(Flat): use global intelligence only with single interest
    • SOS [1]

Evaluation of social intelligence

- **Accuracy**
  - Largest precision rate: quality
  - High recall rate: completeness

- **Efficiency**
  - Largest response rate: incentive
  - Short response delay: time efficiency
Evaluation of global intelligence

- **Accuracy**
  - Largest precision rate: quality
  - Largest recall rate: completeness

- **Efficiency**
  - Largest response rate: incentive
  - Comparable short response delay: time efficient
Outline

• Introduction
• Related work
• iASK design
• iASK implementation
• Evaluation
• Conclusion
Conclusion

• iASK: a unified distributed Q&A system incorporating both social community intelligence and global collective intelligence
  ◦ A neural network to consider multiple factors in evaluating the answer QoS of a user’s friends
  ◦ A virtual server tree overlay to efficiently locate answerer candidates in the interest of the question
  ◦ A fine-grained reputation system to locate cooperative global experts

• Future work:
  ◦ Add more features to rank users in order to more precisely and efficiently locate the experts
Thank you!

Questions & Comments?

Haiying Shen
shenh@clemson.edu

Electrical and Computer Engineering
Clemson University