User Behavior in Q&A community: an analysis of Zhihu

Andi Liao 2018/04/04

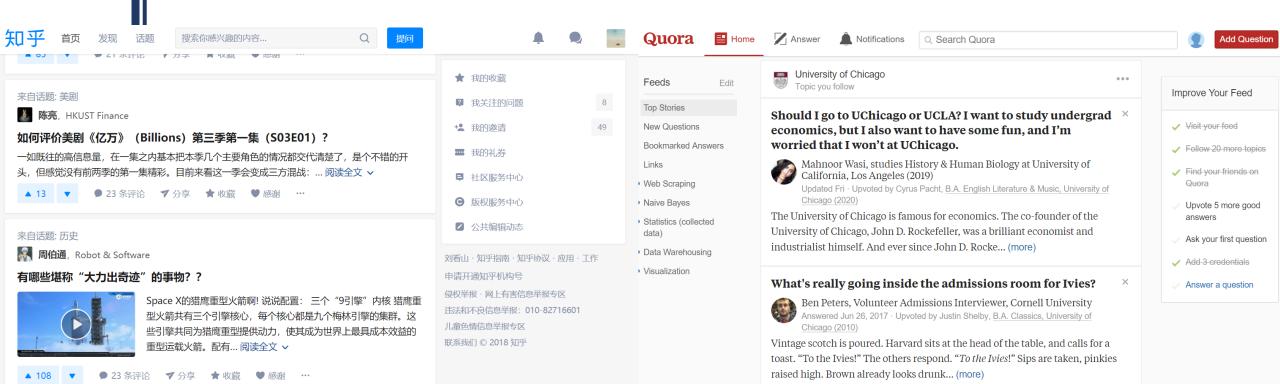


Brief Intro to Zhihu

- Similar with Quora
 - Q&A Community
- Follower & Following, Upvotes & Downvotes

User Interface

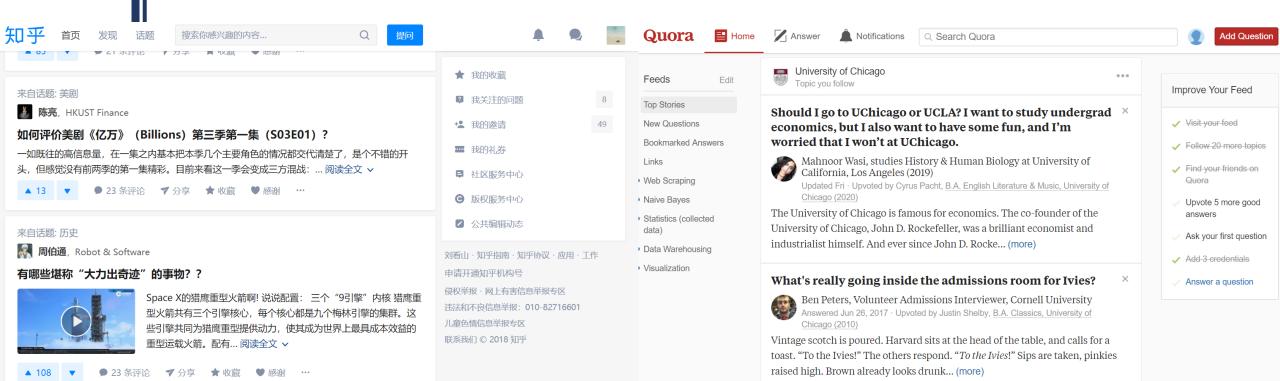
- Navigate Bar, Topics
- Recommend System Feed, Invitation





Brief Intro to Zhihu

- Difference with Quora:
 - User Identity Any Name
 - Upvote Agree
 - Top Writer Centered





Research Question



User Behavior in Zhihu Community

- Classification:
 - Is it possible to <u>cluster</u> users based on data of their activities?
- Prediction:
 - Is it possible to predict a new user to be a top writer based on data of his activities?





Why It Matters?



Really Interesting!

Mixture: Q&A + social

Venmo: Finance + Social

• Weibo: Nickname + Comment

Top Writers preferred?

• Top writers == Users who answered a lot of questions?

Top writers == Users who has many followers?



Why It Matters?



Possible Contribution

- Many scrapers, but few analysis
 - Mostly focused on "questions", rather than users
 - Descriptive statistics
- Help improve the community quality
 - Top writers & Top answers
 - Recommended contents



Data Source

- Scraper: https://github.com/MatrixSeven/ZhihuSpider Java
 - Backup: https://github.com/7sDream/zhihu-oauth Python unofficial API

Data form: Mysql database

Follower	User	User Information
user_name	user_id	company, job
follower_name	index_url	education
update_time	token	answer, question
•••	•••	•••



Possible Theory

- Social and Interaction graph for social computing system:
 - Degree-distribution
 - Clustering coefficient
 - Reciprocity and balance
 - Assortativity
 - Tie Strength
- Reference:
 - Zhang, X., Tang, S., Zhao, Y., Wang, G., Zheng, H., & Zhao, B. Y. (2017). Cold Hard E-Cash: Friends and Vendors in the Venmo Digital Payments System. In *ICWSM* (pp. 387-396).
 - Wang, T., Chen, Y., Wang, Y., Wang, B., Wang, G., Li, X., ... & Zhao, B. Y. (2016). The power of comments: fostering social interactions in microblog networks. *Frontiers of Computer Science*, *10*(5), 889-907.
 - Wang, G., Gill, K., Mohanlal, M., Zheng, H., & Zhao, B. Y. (2013, May). Wisdom in the social crowd: an analysis of quora. In Proceedings of the 22nd international conference on World Wide Web (pp. 1341-1352). ACM.



Method & Tool

- Classification & Prediction
 - Random forest classifiers
 - Challenge:
 - Supervised learning methods
 - Categorical variables exist
- Reference
 - Patil, S., & Lee, K. (2016). Detecting experts on Quora: by their activity, quality of answers, linguistic characteristics and temporal behaviors. Social network analysis and mining, 6(1), 5.
 - Grosseck, G. (2012, January). An empirical analysis of the educational effects of social media in universities and colleges. In The International Scientific Conference eLearning and Software for Education (Vol. 1, p. 167). "Carol I" National Defence University.



Expectation











Non-Social

Social