A REPORT ON

Signals Matter: Understanding Popularity and Impact of Users on Stack Overflow

Abstract

Stack Overflow, a Q&A site on programming, awards reputation points and badges (game elements) to users on performing various actions. This paper investigates the role of these game elements, namely points and badges in judging 2 main social qualities of the user which are popularity and impact. This system of virtual rewards incentivizes people to contribute in any way possible. It raises the question that whether these elements can tell us something about the user's social qualities, and if they can, are they correct according to the user?

Summary

On investigating the value of these game elements, a problem we face is *Adverse Selection* which creates an imbalance in participation, and for that *Digital Signaling* is used. A signal can be any sort of indication which allows the users to communicate with other users. Signals that are difficult to produce are usually the ones that carry more value which is seen usually in the form of answer badges. The platform is *gamified* through these signals where every element has a specific value attached to it. The survey conducted on the users gave mixed results out of which two possible hypotheses were generated which looks for correlation between these elements and social attributes of a user and which game element is a better indicator for identifying these attributes.

To operationalize these metrics of popularity and impact, two variables *Popularity score* and *Impact Score* are defined which capture all of the user's activity and contribution on the platform and therefore are suitable metrics to verify the hypothesis. On putting all these features in a non-linear ML model, we conclude that Badges capture a more nuanced summary of the user's contribution and it takes into account other factors except upvotes and acceptance. All this might not be captured properly while maintaining reputation scores. Therefore the hypothesis that was made earlier is nullified as we find badges to be better indicators of popularity and impact.

Inferences:

- Answers contribute more to a user's popularity and impact than questions.
- Badges are better indicators of a user's Impact instead of reputation points.
- Non- trivial badges and reputation points are correlated positively with Popularity and Impact.
- Badges capture a more accurate value of a user to the platform.
- People with high Popularity and high Impact have distinct kinds of engagements to the platform. Popular users and more active whereas users with high impact have more answer badges which are difficult to earn. Most of the users have low Popularity and low Impact

This paper draws a very clear line between the qualitative and quantitative qualities of a user which is a very important distinction many people using the platform fail to catch. It shows that answering questions is the best way to gain prominence on the platform. It also helps other users identify more accurately, the people they follow on the platform based on these social attributes which is one of its major unique contributions to the community.

Many people who come to the platform mostly post questions and look for answers and might not even acknowledge the writer for that answer. Therefore profile visits are not a very good signal for maintaining one's reputation here. This also creates a bias for the new users as it is an absolute measure.
It is better if the views on a profile and the number of badges of a user are normalized with his/her age on the platform. Also, Question Badges that are awarded by the users may be based on their personal preferences of just 1 user which leaves scope for human error. Hence they should not be considered while making the model. Similarly, participation badges and many others neither relate to Impact nor Popularity and hence should also not be considered.

While the current model works nicely, some more work can be done on it/ extended further to help the SO community even more. Another factor, the "correctness" of the user can also be added to the platform. We already saw that answer badges are the most important factors in determining one's impact/popularity. In my opinion, it is also important to ensure that the user first makes sure that their answer is correct instead of carelessly posting it in hope of badges and points and misguiding other users. Such users should be penalized and hence, a factor *percentage correctness* can be introduced which maintains the user's accepted/not-accepted answer ratio. As impact is highly correlated with the number of badges, and no badges are awarded when an answer downvoted multiple times, this scenario fails to capture the negative impact of the user. As the negative score of a user only reduces his reputation score which cannot fall below one, the negative impact goes unseen. This idea can be used to identify malware propagators/people with a negative impact which will help users avoid them which could be a significant extension of the project and will help the community a lot.