honours project

1 Principal Goal:

Produce a learning algorithm that determines how ThisThat public polls are distributed among users.

This project has been proposed by Max Osborne (osborne@thisthatapp.com). This is an industry supported project, and may be subject to intellectual property assignment. Contact the proposer for more details. You have been approved but not selected for this project.

ThisThat is a smartphone app based polling platform, where users can ask and answer polls, participate in debates and gather statistics on binary questions.

It is the student's task to create an algorithm that prioritises the order in which polls are shown in each user's Flow page (the page where public polls are visible in a card slider format), where prioritisation is measured with a prediction of highest user engagement.

Predicting user engagement is a hard problem and while optimising the distribution could be done exactly through co-operative game-theory, such work would be intractable and thus the student is recommended to look in to utilising heuristics. ThisThat have a proposed approach and recommend the student considers this.

ThisThat's proposed approach: Make the core of this project determining unique user coordinates in a multidimensional space that is created from ThisThat's database attributes, in which clusters of identifiable traits should be formed. These user coordinates provide a similarity distance measure of users. Given that polls are created by users, we can infer a measure of engagement of a poll through users' relative distances, and hence respectively determine broadcasting speeds of polls'.

The process of obtaining user coordinates is recommended to go as follows: First define a space from the normalised attribute data of ThisThat's database and for efficient computation reduce the dimension of the large matrix containing this data using some order reduction method (e.g. PCA). Then create clusters by grouping the users in a manner that attempts to identify some relationship between them and find the optimal coordinates for the cluster centres. As new data becomes available, user coordinates should be updated and as a consequence the cluster centres should be adjusted accordingly.

2 Possible extension(s):

1. The student can consider a multiplier feature emulating tinder's boost feature, where a higher multiplier results in a quicker response collection from the public for a given user's own publicly posted polls. In other words this multiplier could multiply the speed at which their polls are distributed to other users. Such a multiplier value would sequentially be updated based on the users' response rate to the public polls and thereby incentivise them to answer questions from the public.
2. The student can help ThisThat's mission of bridging society by minimising groupthink. While user-coordinates (as mentioned above) reveal similarity in users and hence help tailor distribution of polls to increase engagement, these co-ordinates may also reveal homophilies of opinions. Indeed interpreting the distance from user to user as the strength of users' links, where users are the nodes, we can identify

clusters of similar opinion and weak ties for groups of different opinion. We'd like to create more links between such weak ties to diversify knowledge and opinions and see an opportunity in doing so by taking inspiration from exploration vs exploitation in Reinforcement Learning literature. Providing the user with the ability to choose a percentage of how tailored their polls are to them and translate this to the ratio at which we exploit (broadcast polls at a speed relative to user-coordinates) or explore (randomly broadcast polls). The aim is that users hereby engage in conversation with polls coming from weak ties.

3 Difficulty 4 - Very Hard

References

1. Clustering of users in social networks by their activity (http://dx.doi.org/10.1109/ukrcon.2017.8100415)
2. http://www.scitepress.org/Papers/2017/62028/62028.pdf
3. https://arxiv.org/pdf/1505.02399v2.pdf