

Proposal: Analyzing social interactions and relationships in OSN-based Gifting applications

Zejun Huang, Hong Fan, Xu Han
zejhuang, honfan, xuhan @ucdavis.edu

1. Data Description

The data is collected from 2 facebook gifting application: ismile and iheart. And we also clawed the user feature data including gender and locale from Facebook. The raw data contains: sender_id, receiver_id, time_stamp and gift type. The original data contains 2 years' of user behavior data.

We picked two weeks (2009-50, 2009-51), which are the peak active weeks for application ismile and iheart. It would be interesting to compare these two applications in the same time period and they are also at the same phase of evolution.

We built the directed graph based on behavior of gifting, i.e. userA sent a gift to userB 3 times, then there is an edge from userA to userB which has the weight of 3. In the ismile graph, we have 16,231,301 nodes, 20,446,940 edges and average degree is 2.519; in the iheart graph, we have 62,752,508 nodes, 119,896,569 edges and the average degree is 3.821. We should notice that according to this simple statistic, iheart is more popular than ismile.

We also have the gender and locale features for the users, and we can match users in different application since they have the same user id.

2. Research Problems

We would like to analyze social interactions in four perspectives.

First, investigating social interaction styles of individuals. These styles include interactive, immersed or isolated. Also, based on degrees' in the graph, we can identify users who are active or passive.

Second, revealing social relationships between two individuals. We want to identify the edges which are reciprocal (bidirectional) or parasocial (unidirectional). We will also compare the features for the users in the same link, and to see if the type of link is affected by user behavior or by domestic features (gender, locale). We will also compare the neighbor similarity for the two individuals.

Third, cross graph analysis. Since we have graph built on the same phase and time of the application, we can compare users' behavior in two applications. The number of common users is about 886 thousand. We would like to see if one's behavior type (active, passive, interactive, etc.) would stay the same in the 2 different applications. We will also compare the neighbor similarity in the 2 graphs for the common users.

Last, prediction. We want to utilize the information we have above to predict the link type in a new graph (application). If we can find the same behavior pattern in step 3, then it is reasonable to use the features of ismile to predict the behavior and link type in iheart.

3. Methods

3.1 Techniques and Softwares

We first need to calculate the features for user (node) and links (edges), this is required to use python and SQL. All data are nearly prepared in postgresQL. Then we will use R for visualization and regression analysis. For the prediction part, we are considering using 'scikit-learn' in python, which is a very powerful package in machine learning. Simple SVM might be suitable for our data, since at last we are going to predict a simple 'yes/no' question.

3.2 Potential Problems

The problem with the graph as we can see it now is that it has many leaf nodes, which means that these nodes only received gifts but don't respond. We still want to keep the data unbiased by keeping those nodes, but the amount of the these nodes (more than 60%) may have a big influence in our data.

Also, since we have a graph with more than 1 million nodes, it is difficult to run graph algorithm on it. It would be interesting to detect community in our graph, however the size of data will definitely limit our choice.