Northwestern University,
Paper Summary
Yuxuan He
NetID: yhh325

A summary of "Everyone's an Influencer: Quantifying Influence on Twitter "

by Yuxuan He

Introduction:

In the article "Everyone's an Influencer: Quantifying Influence on Twitter" by Eytan, Jake, Winter and Duncan, we learned that "ordinary influencer" who are exert average or even less-than-average influence could be used to obtain the most cost-effective performance. The authors conducted several experiments, they obtain a predictive model of cascade size based on Twitter's data to test the influence of specific individuals to seed content.

Summary:

The study investigated 90 million Twitter users and tracked about 74 million diffusion events over two months. Then the authors put the followers of the users into a queue to be crawled, I think we could regards the users and events as a pair of URL(shortened as bit.ly) and seed. The article also showed us the statistics of in-degree(friends) and out-degree(followers), even the number of seeds.

In order to calculate the influence score of each URL post when a user post a URL after two other neighbors already posted it. The author described three method to assign the influence to multiple source. The first method is first publisher got the credits so called first influence. The second method is most recently publisher take the credit, defined as last influence. The third method is split credit, it means the credits is divided equally by the publishers.

The author noted that the influence of URL is more inclusive than "retweet", so the estimate of influence could be viewed as an upper bound. On the other hand, the author also said they might underestimate the actual influence, so they relied on reposting as a conservative measure of influence. The distribution of cascade size seems to be followed as power-law, but the distribution of cascade depth seems to follow exponential-law.

The author used a "regress tree model" to model the influence. The regression tree model is better than OLR model because it can integrate different simple predictor into a tree, so that it is easy to handle cascades with varied size. The model includes serval user attributes such as # followers, #friends, # tweets and date of Joining. It also includes serval past influence such as "avg, min, and max total influence" and "avg, min, and max local (immediate followers) influence."

Moreover, the authors study the effect of content over the diffusion of information. They also found that URLs classified as interesting and associated to positive feelings tend do generate larger cascades.

Improvement:

The authors try to find a solution to a challenging question—"If you are a marketing professional and want to hire some twitters to spread your content, which users would you choose?" Two obvious solution is find a lot of "low rank" or "low star" users to spread your stuff. However, the author of this article think there would exist a balance between these two solutions. If you find a optimal point, the results of calculation shows it would be a good strategy for marketing professional.

Brief assessment

The article shows good model to predict the influence on twitter, it could clearly show us the results of "ordinal user vs influentials". But I am a bit curious about would the model still work if the size of the datasets become larger. Would it be better if it applied by other algorithms, I mean we could try some other model such as GA. After all, the article is really good for us to know the influence of social network.