Behavior Analysis in Social Media

Reza Zafarani and Huan Liu, *Arizona State University*

With the rise of social media, information sharing has been democratized. As a result, users are given opportunities to exhibit different behaviors such as sharing, posting, liking, commenting, and befriending conveniently and on a daily basis. By analyzing behaviors observed on social media, we can categorize these behaviors into *individual* and *collective* behavior. Individual behavior is exhibited by a single user, whereas collective behavior is observed when a group of users behave together. For instance, users using the same hashtag on Twitter or migrating to another social media site are examples of collective behavior. User activities on social media generate behavioral data, which is massive, expansive, and indicative of user preferences, interests, opinions, and relationships. This behavioral data provides a new lens through which we can observe and analyze individual and collective behaviors of users.

The emergence of this new type of data presents behavior analysis on social media with new challenges. We detail first what individual and collective behavior analysis is, and then outline novel challenges with future work.

Individual Behavior Analysis

Individual behavior can be considered one of the following:

1. *User-User Behavior.* This type of behavior is observed between two users. For example, befriending and following in social media are examples of such behavior.

2. *User-Entity Behavior.* This type of behavior is exhibited with respect to entities on social media (for example, user-generated content). For instance, liking a post on Facebook or posting a tweet on Twitter are examples of user-entity behavior.

3. *User-Community Behavior.* This is the type of behavior that users exhibit with respect to communities. Joining and leaving communities are examples of this type of behavior.

Irrespective of the type of behavior observed, we can utilize a computational methodology to analyze behavior and help find interesting patterns as follows [5, 4]:

To analyze individual behavior, we can collect traces of the behavior on social media. For instance, for analyzing the user-befriending behavior, we can collect the list of individuals that users befriend over time. To understand the underlying reasons for befriending, we can follow a machine learning tradition and create data features that are likely to be related to the behavior. For the befriending behavior, we can consider the number of common friends as an important feature. After constructing the features, we can find the correlation between features and behavior using a supervised learning framework. Since correlation doesn't imply causation, evaluation techniques are required to verify the validity of our findings. We can use randomization tests [5] or causality testing techniques, such as Granger's causality [5] for evaluation purposes.

Collective Behavior Analysis

Collective behavior analysis can be performed by either analyzing the individuals that exhibit the collective behavior independently, or by analyzing the individuals that exhibit the collective behavior as one group. In the former, we're aggregating (summing, averaging, taking majority, and so on) the result of individual behavior analysis, which can be performed using the aforementioned methodology. In the latter, we consider the individuals exhibiting the collective behavior as one (large) group, and the behavior is analyzed for this group. As the focus is on group-level behavior, we can use methods that model group-level dynamics to analyze collective behavior. For instance, epidemic models from epidemiology or techniques that analyze influence on implicit networks can be used to analyze collective behavior [5].

Case Studies

Let's discuss four behavior analysis case studies. Two are examples of individual behavior analysis (*community membership behavior* and *connecting users across sites*) and the other two (*movie revenue prediction using Twitter* and *user migration in social media*) are instances of collective behavior analysis.

Community Membership Behavior

Users often join communities for different reasons. To analyze this behavior, Lars Backstrom and his colleagues [2] collected information about users joining communities over time and designed features that could have influenced users joining communities. They determined how important these features are in predicting whether users join communities by using a decision-tree learning algorithm. Their findings suggest that not only is it more likely for individuals to join communities when they have many friends inside the community, but it's also important how these friends are connected inside these communities for example, how dense their friendship network is.

Connecting Users across Sites

In previous work [6], we connected users across social media sites by identifying multiples sources of information that are generated by the same user. We noticed that the minimum information available on different social media sites is the username individuals select. By employing usernames alone, we identified profiles that represent the same individuals across social media sites. We analyzed behaviors that individuals exhibit when selecting usernames, such as selecting the same usernames, using the same language or vocabulary, and their typing patterns, among other behaviors. These behaviors were captured using data features and helped successfully connect users across sites.

Movie Revenue Prediction Using Twitter

Sitaram Asur and Bernardo Huberman [1] attempted to predict the collective behavior of going to the movies by analyzing the traces it left in the microblogging site Twitter. They found that by employing only eight features, movie revenue can be predicted with high accuracy. These features are the average hourly number of tweets related to the movie for each of the seven days prior to the movie opening (seven features) and the number of opening theaters for the movie (one feature).

User Migration in Social Media

Working with Shamanth Kumar [3], we analyzed the collective behavior of users migrating across sites. We showed that using three general features that measure user's activity, user's network size, and user's prestige, we can effectively model and predict populations that migrate across social media sites.

Behavior Analysis Challenges

User behavior analysis in social media faces stern challenges. Here, we outline some immediate and demanding issues.

Data Sparsity

Not all behaviors are easily observable on social media. Consider analyzing the money individuals spend on social media or their driving routes. These data aren't as abundantly available on social media as they are in the physical world. In other words, while for specific patterns (such as befriending) massive sources of data are available on social media, for other behaviors data are sparse. This imbalance in data availability limits the behaviors that can be analyzed using social media and at the same time provides opportunities for identifying relevant information sources for behavior analysis.

Lack of Causality Information

Behaviors in social media are only observed by the traces they leave in social media. We rarely observe the driving factors that cause these behaviors; nor can we interview individuals regarding their behaviors. Consider a tweet of John that is retweeted on Twitter by Mary. Does that mean that the tweet is interesting to Mary? Does that mean that the tweet is worth spreading to others? Or, does that mean that John and Mary are friends and Mary retweets all tweets by John? These interesting questions are frequently encountered when trying to identify causes of behavior in social media.

Evaluation Dilemma

Even if a behavior is analyzed on social media and related patterns are gleaned, it's difficult to verify the validity of these behavioral patterns. Evaluation becomes even more challenging for industries in which important decisions are to be made based on observations of individual behavior. For instance, consider an online seller that observes an abnormal rise in their site visits and sales. The site can expand and invest in the infrastructure to handle such traffic. However, this traffic could be due to a sudden demand; hence, temporary. A deeper understanding of its members is necessary to ensure the validity of these behavioral patterns.

The journey of understanding human behavior at scale has just begun. Most current work in social media considers analyzing behavior from a data mining or machine learning perspective. Social sciences, including psychology, anthropology, and ethnography, have developed their theories and rigorous methods to understand human behaviors in small groups and at a small scale. It's imperative to generalize these theories to larger populations observed in social media and to design new techniques tailored to behavioral analysis on social media.

References

[1] Asur, S., Huberman, B. A. (2010, August). Predicting the future with social media. In Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on (Vol. 1, pp. 492-499). IEEE.

[2] Backstrom, L., Huttenlocher, D., Kleinberg, J., Lan, X. (2006, August). Group formation in large social networks: membership, growth, and evolution. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 44-54). ACM.

[3] Kumar, S., Zafarani, R., Liu, H. (2011, April). Understanding User Migration Patterns in Social Media. Proc. AAAI, 2011.

[4] Longbing Cao, In-depth Behavior Understanding and Use: the Behavior Informatics Approach, Information Science, 180(17); 3067-3085, 2010.

[5] Zafarani, R., Abbasi, MA., Liu, H., Social Media Mining: An Introduction, Cambridge University Press, 2014.

[6] Zafarani, R., Liu, H. Connecting Users across Social Media Sites: A Behavioral-Modeling Approach, In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 41-49). ACM.