Some Research Opportunities on Twitter Advertising

Milad Eftekhar
Department of Computer Science
University of Toronto
Toronto, ON, Canada
milad@cs.toronto.edu

Nick Koudas
Department of Computer Science
University of Toronto
Toronto, ON, Canada
koudas@cs.toronto.edu

Abstract

Social media have enjoyed a rapidly increasing adoption among users in recent years. Millions of users execute billions of actions (in form of tweets, messages, replies, likes, etc.) every day. The massive amount of social interactions online has contributed to the proliferation of social advertising.

Alongside the interest in online and social advertising, Twitter has introduced several advertising opportunities to aid advertisers to promote their products/services to the "right" audiences. This includes providing advertisers with (1) different advertising options of "Promoted Tweets", "Promoted Accounts", and "Promoted Trends", and (2) different user targeting options based on keywords, interests, location, etc. In this paper, first we briefly discuss the Twitter advertising platform, and then we introduce some research problems in this domain.

1 Introduction

Social media is used daily by millions of people (including but not limited to, journalists, celebrities, business owners, charities, etc.). Services such as Twitter, Facebook, LinkedIn, and Pinterest allow millions of users worldwide to interact with each other, share and consume content.

Micro-blogging platforms such as Twitter have experienced significant growth in user acquisition and participation in recent years. Twitter enjoys worldwide adoption with 500M registered users who generate over 400M tweets and 1.6B search queries every day. Social connections are established by "following" users. When a user u follows a user v, u sees all tweets posted by v in its timeline. The timeline of a user (say u) is the set of tweets generated by any user that u follows.

Given the wide user success Twitter enjoys, the focus on monetization for the platform has been on advertising and marketing. Users utilize Twitter as a marketing tool to broadcast messages to their followers. Companies, organizations, celebrities, etc. take advantage of this opportunity to target their followers for different purposes (e.g., brand/product awareness, sales leads, or general information dissemination, etc.) by broadcasting messages that will appear in followers' timelines.

Several companies including Google enjoy wide success based on a keyword based advertising model. Users utilize the search engine for search and business bid on the search keywords, taking the opportunity to showcase ads in suitable places of the screen along with the search results. In a social setting however, as is the case of Twitter, keyword search is only one functionality aspect. Twitter users follow other users

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

consuming content. The opportunity to mix content consumption and advertising is unique in a social setting. Typically the content consumed by an individual is highly specific; understanding the type of content each user consumes offer great opportunities for highly tailored advertising.

Recently Twitter introduced several advertising models ranging from the typical keyword search model (that is very common in many online advertising platforms such as Google AdWords) to advertising models based on interest targeting. In Section 2, we take a quick look into the different advertising models introduced by Twitter. Considering the interest-based targeting models, there exists a clear need to identify the interests and expertise of users in a social platform. Section 3 details how we can locate different topics of expertise and interests for each user utilizing the Peckalytics system [1], followed by Section 4 introducing research problems in this domain.

2 The Twitter Advertising Platform

Twitter recently introduced advertising models aiming to fulfill diverse marketing and advertising functions on the platform. These models can be categorized into two main groups: (1) Models that facilitate advertisers to better promote their products/services, (2) Models that aid advertisers target specific users more effectively.

Advertisers typically promote three types of products: tweets, accounts, and trends. With the "promoted tweets" option, a tweet is provided by the advertiser and it is inserted in the timelines or search results of all users who are targeted. The "promoted accounts" option enables advertisers to acquire more followers and build a larger community of advocates. The promoted account (provided by advertisers) is shown in the search results and in the "Who to follow" section of targeted users' profiles. "Who to follow" is a section inserted next to any user u's timeline suggesting what users u is likely to be interested in following. The results in the "Who to follow" section are generated by Twitter's recommendation engine [4]. Finally, advertisers may sponsor trends utilizing the "promoted trends" Twitter advertising option. Trends are hot topics of the day and are placed next to the timeline of each user. Trends can be promoted globally or regionally. Consequently promoted trends offer wide exposure.

Twitter also introduced several advertising options to target users. Besides the keyword targeting model that is popular in search engine advertising (e.g., in Google AdWords), Twitter introduced a new advertising model that allows advertisers to target users by their interests, geography, and gender. Moreover, followers of a set of Twitter accounts can be targeted as well as Twitter users who are similar (have similar interests) to these followers (or to any specified set of users) [6].

When targeting users based on interests, an advertiser is able to provide a promoted tweet or a promoted account and target it to a set of users who have specific interests. Interests are explicitly chosen by the advertiser. In particular, an advertiser chooses a topic (or set of topics) t and provides a promoted tweet or a promoted account. Based on proprietary algorithms, the Twitter advertising platform identifies all the users who are *interested* in the topic t; the promoted tweet is inserted in the timeline or search results of these users (explicitly identifying it as a promoted tweet); or suggests these users to follow the promoted account. For example, one can conduct an advertising campaign on a wine festival providing a tweet and selecting topics such as "wine", "tourism", etc. The provided tweet will be inserted in the timelines and search results of users who are interested in "wine" and/or "tourism".

Users can also be targeted by geography. A promoted tweet or account is promoted to a group of people who live in a specific location (country or metropolitan area). These options aid advertisers to target regional audiences. Naturally regional targeting is preferable when the event or product is of regional interest only.

Advertisers also have the option (a) to provide Twitter accounts and target a promoted tweet to the followers of these accounts and (b) to target users who are similar to the followers of some accounts. Assume one aims to advertise a new independent movie. Naturally it makes sense to target those Twitter users that are interested in independent movies. Notice that those interested in independent movies may or may not tweet about such movies. Primarily they will be consuming content from Twitter users that are experts in independent movies and produce content primarily about such movies. Utilizing option (a), one can identify the "*experts*" on " independent movies" and target the followers of these experts. By following accounts who have expertise on a topic such as independent movies, these users show interest on the topic; hence, they are good targets for promoting a tweet on the new movie. Utilizing option (b), we look for users who are similar to (have similar interests) the followers of experts on independent movies. These include all users interested in independent movies.

Central to these approaches is the ability to identify "expert" users on particular topics on Twitter. In Section 3, we explain some approaches including that of Peckalytics that utilizes crowd-sourced information and social connections to identify the expertise and interest of users.

3 Identifying Expertise and Interests

It is expected that users generate content on topics they know well or care about. The expertise area of user v is the set of topics that v knows well, or is known for in the community. For example, a company shares tweets related to its products whereas a soccer player post tweets about the team, the upcoming events, some personal life information, etc. Technically, by sharing information that is related to one's expertise area, one aims to become more popular to users who are interested in that field. This leads to gain more followers, attention, and observability. On the other hand, one can understand what a user is interested in, by examining who they follow. When user u follows user v, u expresses interest in the content v generates.

Let $U = \{u_1, u_2, \cdots, u_n\}$ be the set of users and $T = \{t_1, t_2, \cdots, t_m\}$ be the set of topics. For a user u, let T_u be the set of topics associated with u. It is feasible to obtain the set of topics mostly associated with a Twitter account. These are usually the topics of expertise for the account holder. Several approaches could be applied to assign an account to a list of topics declaring its expertise. Machine learning technology is mature enough to classify the tweets one generates based on topics. Alternatively, the approach taken by the Peckalytics project [1] provides results taking a crowd sourcing approach.

Definition 1: A user $u \in U$ is an *expert* on topic $t \in T$, iff $t \in T_u$. This means that (for our specific way of extracting topics) other Twitter users recognize u as an expert on topic t. We call topic t, a topic of expertise for u.

Assume $C = \{c_1, c_2, \cdots, c_r\}$ is the set of contents generated by users in U. In the Twitter setting, a content can be a tweet, a shared video, a posted link, etc. Moreover, assume there exist a mapping $M : C \rightarrow 2^T$ that maps each content in C to a subset of topics T (topics that are associated with this content). For example, $M(c_1) = \{t_1, t_3, t_5\}$ translates to "content c_1 is associated with topics t_1, t_3, and t_5" where content c_1 can be the following tweet "The first lady Michelle Obama makes surprises by reading the best picture in Oscars 2013" and topics t_1, t_3, and t_5 can be Politics, Hollywood, and Art.

Definition 2: A user $u \in U$ is *interested* in topic $t \in T$ iff the probability that u follows (reads) any content c that is associated with topic t ($t \in M(c)$) is higher than a given threshold $\theta \in [0, 1]$.

The approach taken by the Peckalytics project is to utilize Twitter lists to extract expertise and interests for Twitter users. Twitter introduced the concept of *Twitter lists*. A Twitter list is a collection of Twitter accounts (users). Typically, users create lists annotated with a descriptive name and place their favorite accounts who are perceived by the creator of the list to be experts on a particular topic (typically the list name) into the list. For example, a user may create a list with the name "politics" that includes Twitter accounts @BarakObama, @AngelaMerkel, @HillaryClinton, @JohnKerry, and @DavidCameron. Lists facilitate content filtering by topic. In other words, by creating a list on "politics", a user can filter the tweets in the timeline to see just the tweets generated by accounts in that list. Creating lists on different topics by different users is a typical experience in Twitter. A user can create multiple lists and a Twitter account can belong to any number of lists.

Recent works have utilized Twitter lists to address some problems such as identifying users' topics of expertise [1, 2, 5], separating elite users (e.g., celebrities) from ordinary users [7], and sampling user-generated data in social networks utilizing an expert based method [3].

Peckalytics utilizes the Twitter lists to identify topics of expertise and interest for different users. It crawls the lists and utilizes Apache Lucene to store and index the lists and the users in the lists. This process is executed constantly as new lists are regularly added and existing lists are dynamically modified by the creators. Peckalytics associates with each Twitter account, a set of topics extracted from the names of the lists containing that account. The process of extraction includes tokenization of the name, common word (stop word) and spam filtering, entity extraction, and related word grouping via Wikipedia and WordNet. An index mapping each Twitter account to the set of topics associated with the account is constructed and managed by Lucene. Thus, for each Twitter account u, a set of topics that best describe the topics associated (by other Twitter users) with u is identified. We refer to this set of topics as the *expertise vector* of user u.

The Lucene index supports all kinds of typical search query syntax including regular match (match in any order, e.g., *social media*), phrase match (e.g., *"social media"*), boolean expressions (e.g., *social AND media*), negative match (e.g., *-social media*), etc. These matching types are also supported by Twitter advertising platform.

To identify the set of interested users for any topic t, Peckalytics utilizes the experts and the social connections between Twitter users. Peckalytics reports the set of followers of experts on topic t as the set of users who are interested in t. This is due to the assumption that a user u follows another user v provided that u has an interest on the contents v generates. Therefore for each user u, we can identify the set of topics, u has an interest (referred to as the *interest vector* of u) as the union of the expertise vectors of the users u follows.

Peckalytics provides the following main functionality:

1. Identifies the expert Twitter users for any topic t. This functionality is helpful when an advertiser aims to provide a set of Twitter accounts and promote a tweet to the set of followers of these accounts. The experts on t are the most relevant accounts to provide while initiating an advertising campaign on t.

2. It offers analytical functions on the set of experts for any topic t. These functions include the identification of other topics of expertise for a user, their conversations (e.g., the frequent keywords, keyword pairs, and hashtags in their tweets), and the most popular sites they share content from. These analytics aid to assist the selection of topics or keywords an advertiser could choose for campaigns.

4 Some Research Problems

The Twitter advertising platform offers new research problems that could be of interest to business and advertisers. This section introduces three research problems in this domain.

Alternative topics: When targeting a specific set of users, it is natural to ask how can we target them with the lowest cost possible. In particular, assume one aims to advertise on topic t. Clearly, different topics have different costs based on the popularity of the topic, number of advertisers that target the topic, etc. The question is whether we can target the same or approximately the same set of users by advertising on a cheaper topic t' instead of t. We call t' an *alternative* topic.

For example, suppose we want to conduct an advertising campaign on *wine*. Let's say, we understand that most of the users who are interested in *wine* are also interested in *tourism*. Moreover, suppose that the cost of advertising on *wine* is higher than *tourism*. Instead of advertising on *wine*, we can advertise on *tourism*, pay less, and still target a similar audience. In this example, *tourism* is an *alternative* topic for *wine*. Identifying the alternative topics, therefore, is the first problem of interest. Moreover being able to quantify precisely the similarity between the audiences of two alternative topics is of interest as well. It creates an interesting optimization trade-off between similarity of audiences and cost for alternative topics.

This problem can be studied in two scenarios: *(1)* The set of audiences for each topic is unknown, or *(2)* The audience sets are given or can be computed based on the existing information. In both, the problem requires solutions that (approximately) measure the similarity of different topics (based on given or calculated audience sets) for a specific input topic and identify topics with lowest cost.

Combining interests: A second problem is to identify the different interests of a set of users. A set of users interested in topic t (I_t), can be more useful for advertising purposes if we are able to understand other topics users in I_u are interested in. For example, while conducting an advertising campaign on topic *social media*, if we know that users who are interested in *social media* are also interested in *seo* (search engine optimization), we can design a campaign that combines *social media* and *seo* (for example promoting the benefits of social media for SEO campaigns). This could serve us better and attract more attention as the campaign combines different interests of the audience.

One issue in this approach is the cardinality of I_t; potentially the set interests can be large. Conducting campaigns that cover all the topics of interest is clearly impossible. For example assume that users interested in *movies* are also interested in *food, soccer, hockey*. We aim to organize these topics into high-level categories (e.g., by merging *hockey* and *soccer* into a bigger category *sports*) and partition the users in I_t based on these categories. Executing this organization, we can target users in I_{movies} (users who are interested in *movies*) who are interested in *food* with a campaign combining *movies* and *food*; and target users in I_{movies} who are interested in *hockey* and/or *soccer* with a campaign combining *movies* and *sports*.

The goal of the second problem is, therefore, to create a set of users in I_t as input, organize their other interests into high-level categories (e.g., sports), and partition the users in I_t based on these categories. Thus algorithms to partition interests and create high-level categories are required.

Expert refinement: Experts on a topic t may have expertise on other topics too. One problem of interest would be to categorize the experts based on their different topics of expertise. For example, starting with the experts on *cloud computing* we may be able to categorize them into a partition representing the experts on *cloud computing* and *virtualization* and another partition representing *cloud computing* and *data centers*. This organization aids us to have a clearer understanding of different experts. Why is this important? One goal in advertising campaigns is to engage experts into promoting products. Consequently the followers of these experts become aware and possibly adopt the product; subsequently this effect propagates in the network. This behavior is a result of word of mouth in social media. One crucial step to instigate a word of mouth activity is to identify the “right” initial experts to convince (the seeders). Identifying the right initial experts is a challenging task as experts may vary in the presence of different objectives. For example, the set of experts on the partition *cloud computing* and *virtualization* and those in the partition *cloud computing* and *data centers* may each be more suitable while conducting different campaigns. The objective will dictate which one is more suitable in each case.

Computationally, partitioning the experts based on the different topics of their expertise can be a hard task as the “right" number of partitions is not clear. Moreover, it is not clear how topics and experts should be grouped together. This points to the need for soft clustering approaches that optimize criteria specific to the campaign under consideration.

5 Conclusion

The growth in the penetration of social networks in everyday life, opened new opportunities in advertising. Deeper understanding of users based on their actions and disclosed information in social networks, offer opportunities to target users based on their specific interests. In this paper, motivated by the Twitter advertising platform we detailed techniques to identify expert users and users with specific interests on Twitter. Subsequently we presented an initial set of problems that could be of interest to solve in the social (microblog) advertising context. Social advertising is still in its infancy and research in this area could result in real impact.

References

[1] A. Cheng, N. Bansal, and N. Koudas. Peckalytics: Analyzing experts and interests on twitter. SIGMOD Demo Track, pages 973–976, 2013.

[2] S. Ghosh, N. Sharma, F. Benevenuto, N. Ganguly, and K. Gummadi. Cognos: crowdsourcing search for topic experts in microblogs. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 575–590, 2012.

[3] S. Ghosh, M. B. Zafar, P. Bhattacharya, N. Sharma, N. Ganguly, and K. P. Gummadi. On sampling the wisdom of crowds: Random vs. expert sampling of the twitter stream.

[4] P. Gupta, A. Goel, J. Lin, A. Sharma, D. Wang, and R. Zadeh. WTF: the who to follow service at twitter. In *Proceedings of the 22nd international conference on World Wide Web*, WWW '13, pages 505–514, 2013.

[5] N. K. Sharma, S. Ghosh, F. Benevenuto, N. Ganguly, and K. Gummadi. Inferring who-is-who in the twitter social network. In *Proceedings of the 2012 ACM workshop on Workshop on online social networks*, pages 55–60, 2012.

[6] Twitter. Start Advertising | Twitter for Business. `https://business.twitter.com/start-advertising`.

[7] S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts. Who says what to whom on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 705–714, 2011.