Social Listening for Customer Acquisition

Juan DU, Biying TAN, Feida ZHU, and Ee-Peng LIM

School of Information System,
Singapore Management University
{juandu,biying.tan.2012,fdzhu,eplim}@smu.edu.sg

Abstract. Social network analysis has received much attention from Corporations recently. Corporations are trying to utilize social media platforms such as Twitter, Facebook and Sina Weibo to expand their own markets. Our system is an online tool to assist these corporations to 1) find potential customers, and 2) track a list of users by specific events from social networks. We employ both textual and network information, and thus produces a keyword-based relevance score for each user in pre-defined dimensions, which indicates the probability of the adoption to a product. Based on the score and its trend, out tool is able to pick up the potential customers for different kinds of products, such as suits which are daily supplies and diapers which are sudden needs. In order to get a more robust and accurate relevance score, we filter the user network and only consider the off-line close friend network. In addition, we could track users in a more flexible way. Despite of the pre-defined dimensions, our tool is also able to track users by customized events and catch those who mention the event at an early stage.

1 Introduction

According to the statistics provided by the Official Twitter blog[1] on March 21, 2013, the social network Twitter has over 200 million active users creating over 400 million Tweets each day. Social networks has become the main form of media for online users to express their opinions and connect with other users. According to [2], these social media data represents a vault of precious information that could be used by companies for increasing their sales revenue. For example, a user who tweeted about his new born baby on Twitter is more likely to buy diapers. Her friends who saw the posts might purchase diapers as gifts for her as well. This suggests that the tweet's content as well as the social relationships are useful for identifying potential customers. The identification should be early before potential customers purchase the product.

There are some existing tools to expand the markets by social networks such as "Twitter Business"[2], a tool developed by Twitter. However, the relationship they considered is the whole network structure rather than *off-line friends*. We

[1] https://blog.twitter.com/company
[2] https://business.twitter.com/

distinguish our system[3] from the others by focusing on the adoption of product for a user through his off-line close friends. Our system provides the analyses based on two social networks: the English version is based on *Twitter* while the Chinese one is based on *Sina weibo*. Our tool is able to detect more potential customers than other tools as well as providing a feature to track a set of users via specific events found in social media. We employ both *textual information* and *relationships* from social networks and produce a relevance score for each user in pre-defined product dimensions, such as cars with a set of weighted words related to cars. The relevance score depends on the user's volume of discussion on the product such that higher relevance score suggests that the user is more likely to adopt the product. Our system gives the relevance score based on the user's whole tweets to catch some loyal customers. And we also provide trends which shows the evolution and occurrence of events, such as birth of a baby from a user. In addition, we filter the network and puts more focus on the off-line friends by the algorithms proposed in [4]. The relevance scores of a user's close friends will also influence his. Moreover, in order to track the users in a more flexible way, our system is able to follow a list of users and define customized topics. Once the user mentions the topic, he will be picked up.

By an aggregated analysis on the tweets and networks, we are able to find the users who intend to adopt a product. In this paper we give an overview of the architecture of our system and its major modules including a briefing on the algorithms in Section 2. We then demonstrate a case in Section 3.

2 Architecture and Algorithm

Figure 1 shows that our system composes of two main parts. In the back-end part, the tool fetches the social media data and processes the data by calculating the users' relevance score to the product and their off-line friend network. While the front-end provides analysis and visualization based on the scores and networks produced from the back-end. The visualization includes ranking of the user's relevance score, detail profiles of a selected user, off-line social network of a target user, and the event-triggered user tracking. The company could directly market their products to the users by posting tweets to them.

2.1 Data Collection and Processing

There are many social networks like Twitter, Sina Weibo and Foursquare. We first choose one of them as our dataset and collect the timelines, user profiles, and networks by the API they provided.

Relevance Score to Business In Information Retrieval, the basic idea for measuring the relevance of a document to a query has hinged on whether or not a query term is present within a document. A document that mentions a query

[3] `http://research.larc.smu.edu.sg/pa/home.php`

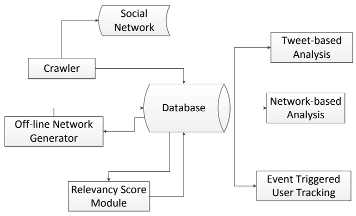

Fig. 1: System overview

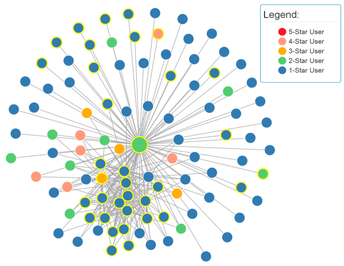

Fig. 2: An example of a user's social network with relevance score

term more often has more to do with that query and therefore should receive a higher score [3]. We adopt this idea to calculate the relevance score of a user to a business.

We define business related dimensions before processing. For each dimension D, we compose a dictionary as $\{t_1 = w_1, t_2 = w_2, ..., t_n = w_n\}$, where t_i is the term and w_i is the corresponding representative weight of t_i. To take the car-insurance as an example, its dimension could be formed as $\{crash = 0.8, gasoline = 0.5, ..., carparking = 0.3\}$. The relevance score $R(tw)$ of a tweet tw to D is determined as

$$R(tw) = \sum_{t_i \in D} w_i * o_i, \tag{1}$$

, where o_i indicates the occurrence of t_i in tw.

And therefore, the relevance score of a user to a dimension D in period $[start, end]$ is produced by $\sum_{tw_i \in [start,end]} R(tw_i)$.

To build up D, we employ the ideas in [1] to extract the keywords automatically. Based on a small set of seed terms t_i which is selected at first, some other representative co-occurring terms are then extracted from tweets. We add the new terms and iterate the selection until the number of terms reach a threshold α. And then, w_i is measured as n_{rt}/n, where n_{rt} is the number of dimension-relevant tweets and n is the total number of tweets containing t_i. The terms with w_i under threshold β will be filtered out.

Off-line Friend Network Since the user will be influenced more by his close friends, we generate the off-line network by the algorithm proposed in [4]. The algorithm utilizes the structural network and applies random walk with restart iteratively to produce the closeness score between users. For the details, please refer to [4].

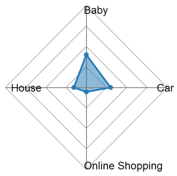

Fig. 3: Radar chart to represent a user's relevance score in different dimensions

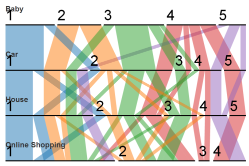

Fig. 4: Joint relevance distribution in different dimensions among user's network

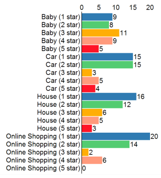

Fig. 5: Relevance distribution in different dimensions among user's network

2.2 Social Network Marketing

Tweet-based Analysis The tweets that the user sent out directly indicates the possibility of the user to buy a product. If a user posted a lot of tweets about fashions, then he is more likely to buy clothes and shoes. For each user, a radar chart for the relevance score on different dimensions is shown as Figure 3. The radar chart illustrates the user's topic distribution in the pre-defined dimensions. The tweets related to the dimensions will also been provided as word cloud. When advertising, the tweets contribute to eliminating the distance between customers and salesmen.

Network-based Analysis Despite of what the user said, the behaviour of other users will also affect his decision to buy a product or not. Just as a famous marketing phenomenon, which is "Since a user care more about what his neighbour thinks than what Google thinks"[2], so the influence of his social network contributes to detecting potential customers. Actually, the influence in the network is two-way. If the target user is a loyal customer, then the neighbors in his network is much easily to be advertised with success. In addition, if the probability of the user's close friends to buy a product is high, and thus the intension of him to purchase becomes higher as well. And consequently, we filter the user's network and only retrieve the off-line friends network. We generate a relevance score for the user by averaging his off-line close friends' scores.

Static and Dynamic Relevance In fact, some topics are long-term like fashion while others are short-term as new born baby. And consequently, the measurement of the user's possibility to purchase the product will be decided accordingly. In order to cope with various products, our system provides not only the user's overall relevance score by all his tweets, but also the score trends by splitting the tweets into different periods. To take the topic "baby" as an example, if the user posted a tweet as *"Listen up. I am PREGNANT."*. Even though the relevance score to the baby dimension is not large since the number of tweets related to baby is small, they will also be found by score trends.

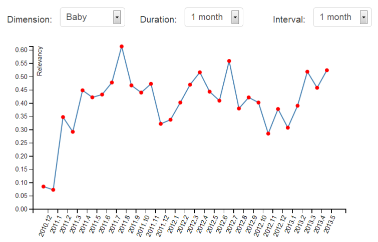

Fig. 6: The user's relevance trends in a dimension

Fig. 7: The user's word cloud in a dimension according to his tweets

Event-triggered Marketing and User tracking A user's relevance score in different dimensions is preprocessed, which could meet the demands of detecting potential customers. However, we still have two problems need to solve.

– How to track the user by customized events?
  We pre-define a weighted word dictionary for each product dimension to produce the relevance score. However, in case of some new events, our system is able to allow system users to customize their own tracking events.
– How to find the user at the early stage?
  Based on the customized events, our system will continuously check whether the tracked users triggers the events. As long as the crawler collects the latest data, once the user triggers the events, he will be highlight immediately and the check frequency is decided accordingly.

3 Demonstration cases

In this section, we demonstrated a case as a baby-insurance salesman. We could search the top baby related users from "Social Network Marketing". Under the tabs "Rank by User Relevancy", "Rank by Relevancy Growth Rate" and "Rank by Close Friend Relevancy", the static individual relevancy, the dynamic individual relevance trend and the dynamic community relevance are shown separately. If we are willing to advertise the insurance, we just need to choose the users and then send out a tweet to them by "Contact Clients". We are able to view the detail profile and the social network of a user as well. The profile page includes user's basic information provided in social networks. Besides, the radar chart for the user in different dimensions are shown as Figure 3. The relevance trend to the baby dimension is represented in Figure 6, where a sudden burst could be found on 2011.02. By viewing the relevant word cloud and tweets as Figure 7 represents, we could eliminate the gaps between the salesman and the user. He could know what the user mentioned in this dimension. Furthermore, in the social network page, the user's 1-hop relationship is shown as Figure 2, where different colors indicate different relevance scores. Red is the highest while blue is the lowest. The yellow bordered dots stand for the user's off-line friends. Analyses for the network such as joint distribution and relevance score distribution in different dimensions are shown in Figure 4 and Figure 5 respectively. In the "Event Trigger Marketing" page, after entering the salesman's login ID "1", we could view his own tracking users. Manually adding some events such as "milk powder" with keywords "kid, milk, powder, buy". Once a follow user mentions the event, he will be highlighted as Figure 8 shows. The refresh frequency could be adjust accordingly.

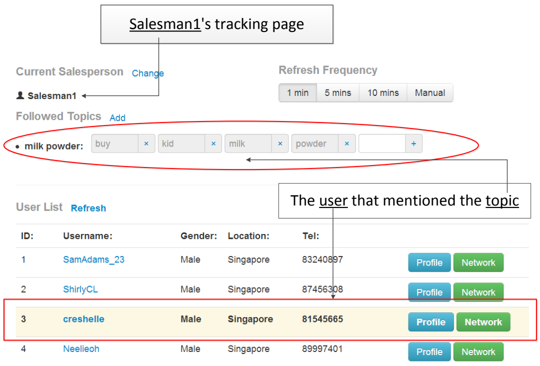

Fig. 8: Tracking users by customized events or topics. (This is just an example and the information like "Tel." is not real.)

4 Conclusions

Our system is an online tool to help corporations to expand the market from social networks. We analysis the users by their own relevance, relevance trends and close friends' relevance score for different types of products. Our system is also able to track users by customized events. In conclusion, we could assist corporations to find and reach potential customers on social networks with ease.

References

[1] Clark, M., Kim, Y., Kruschwitz, U., Song, D., Albakour, D., Dignum, S., Beresi, U.C., Fasli, M., De Roeck, A.: Automatically structuring domain knowledge from text: An overview of current research. Information Processing & Management **48**(3) (2012) 552–568
[2] Kaplan, A.M., Haenlein, M.: Users of the world, unite! the challenges and opportunities of social media. Business horizons **53**(1) (2010) 59–68
[3] Manning, C.D., Raghavan, P., Schütze, H.: Introduction to information retrieval. Volume 1. Cambridge University Press Cambridge (2008)
[4] Xie, W., Li, C., Zhu, F., Lim, E., Gong, X.: When a friend in twitter is a friend in life. In: Proceedings of the 4th International Conference on Web Science., ACM (2012) 493–496