Weibo User Forwarding Behavior Prediction for Reading Promotion

Ma Yue

Shaanxi Normal University, Xi'an, Shaanxi, 710065

Abstract: At a time when the entire Internet is developing at a high speed, especially the rapid development of new media in recent years, the way of information dissemination on the Internet has changed dramatically. Not only has the convergence of all kinds of information become higher, but the amount of all kinds of information has also shown an explosive growth. Hundreds of network users have received information in a very short time. It also allows everyone to participate in the production and dissemination of every piece of information. Nowadays, the biggest way of information dissemination is Weibo forwarding and WeChat forwarding. The forwarding behavior of Weibo users has become one of the fastest factors for information dissemination. The purpose of this paper is to study how to improve the accuracy of prediction of forwarding behavior. According to the characteristics of reading promotion, this paper analyzes the influence factors of the forwarding behavior of Weibo for reading promotion and constructs the corresponding user forwarding behavior prediction algorithm for global prediction.

1. Introduction

Weibo is easy to operate and users can play it freely, making it a place where various of people can get the information they need or share their lives. Sina Weibo, as one of the typical representatives, can provide a very good data base not only because of one of the earliest platforms, but also for the huge group of users. Sina Weibo content can cover a variety of topics. It can be a national event or a trivial matter in life. Now Weibo has become one of the important ways for netizens to obtain information. On the micro-blog Platform, publishing personal information, forwarding information of interest and so on are an important basis for the study of user behavior. Among them, reading promotion can be quickly learned by users, which still needs to be realized through users' continuous forwarding. Through the continuous forwarding of users, micro-blog information can be spread when it explodes. So as to achieve the purpose of popularization. What are the factors that make users forward? We understand from the following research that there will be many factors for users to forward. Perhaps because they support or oppose a certain view. or want to share information they like. The research and prediction of this kind of behavior will help to improve the reliability of reading promotion.

2. Research Background of Micro-blog Users' Forwarding Behavior

According to the data shown in the chart below, the scale of micro-blog users in China was 337 million in the first half of 2018, up 21. 4 million from the end of 2017, accounting for 42. 1 percent of the total online population. The number of micro-blog users in China is growing slowly, which is one of the reasons we need to study. [1]

The table below shows that in the first half of 2018, China's mobile micro-blog users numbered 316 million, up 29. 23 million from the end of 2017, accounting for 32. 6% of the number of mobile internet users. [1]we've found that with the development of smart phones in recent years, more and more of our lives and work are being done on them, making people more and more dependent on them. This makes people more and more dependent on mobile phones. A day without mobile phones can also send people into a state of anxiety.
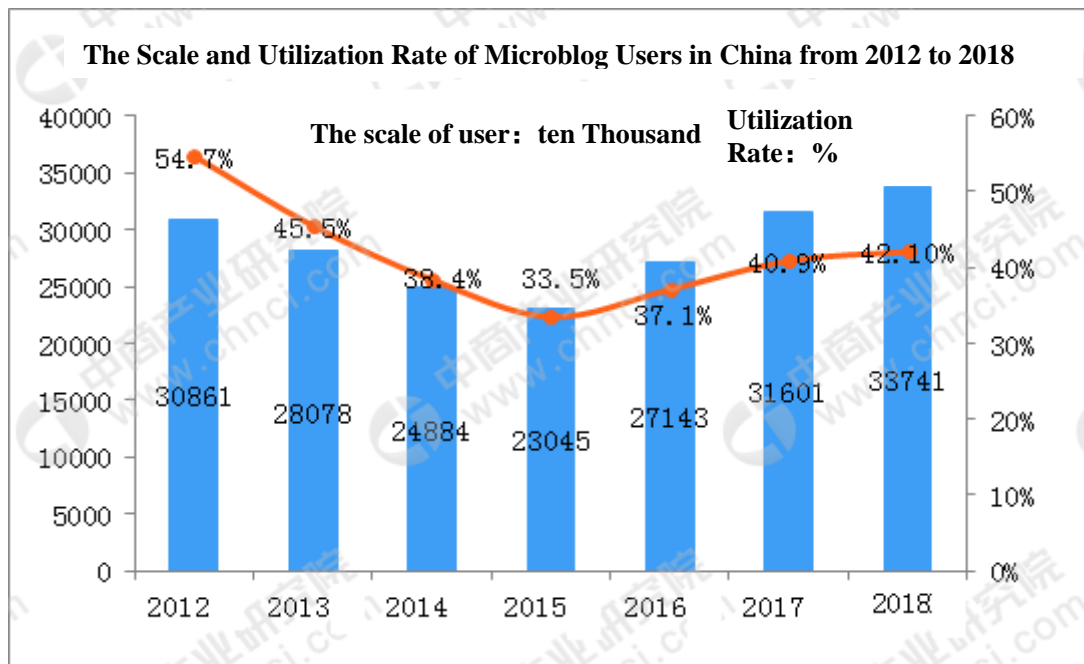
Fig.1 the scale and utilization rate of micro-blog users in China from 2012 to 2018

A large number of users transmit a large amount of information. Although to a large extent to individuals, enterprises and even the community has brought a great deal of convenience, it has also bought many disadvantages, such as information overload, false information flooding and so on. It is these drawbacks that this paper has to face the reading. promotion so as to predict the forwarding behavior of micro-blog users. Based on the existing content and data, this paper hopes to measure the future information transmission trend according to certain methods and laws, so as to predict the user's preference characteristics in advance and to popularize it more pertinently. [2]In order to provide effective anticipation and help for reading promotion, we can study the rules of users' usage time, forwarding content and so on, understand and master the user's preferences. On this basis, the content that the user is interested in can be timely and effectively push to the user. Through the continuous forwarding of users, explosive proliferation can be shown, thus achieving the goal. [3]While making reading promotion more accurate, it also makes the online world cleaner and healthier. At the same time, users can read the information they want in a more comfortable online environment, making the adhesion of the user higher, in favor of their own development.

3. Factors Affecting Micro-blog User Forwarding

We know that everyone is different, so people have different views on an issue. So what do users forwarding for? There are million of reasons for that. We can then translate this forwarding problem into a formal definition. For example, it is known that $p \in P$ represents the publisher of micro-blog, $u \in U$ represents the receiving user of micro-blog, and $v \in V$ represents micro-blog information. Supposing that publisher p publishes micro-blog v at time t, then each micro-blog can be expressed as a quartile of u, v, p. Let ruvp express the forwarding attitude of user u to the micro-blog. User forwarding behavior prediction is to solve ruvp given < u, v, p, t whether core user u will forward micro-blog information v published by user p at time t' (t' > t).

Among them, ruvp can be expressed in many forms. Figure 2 shows the interaction among core users, micro-blogs and micro-blog publishers. As can be seen in the figure, the forwarding behavior of users is the result of many factors. Boyd[4]Boyd et al [5]analyze the forwarding behavior of users on other platforms, then make a summary of the overall, and list the reasons for user forwarding behavior.
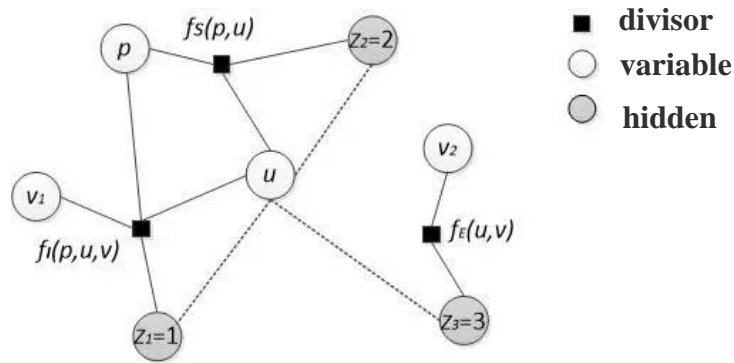
Figure 2. Relationship among Core Users, Microblogs and Microblog Publishers

Through this way and previous studies, we can roughly divide the factors into two directions. The information content is the premise of this study: reading promotion. Whether the reading promotion can be noticed and forwarded by more users, it's all about the content itself and whether the content fits the user's interests: whether the content is rich, whether it fits the current hot spot, whether the user is interested, and so on. And what kind of group users are in is also one of the important reasons that affect users' forwarding. For example, a student will definitely not forward something about teacher reform it is also a very important point what kind of group users are in.

4. Prediction Methods of Micro-blog Users' Forwarding Behavior

4.1. Forecast the past forwarding behavior of users

In general, users will forward topics or content they are interested in, so we can study the user' past forwarding. This will not only be able to understand the user's preferences, but also the user's forwarding behavior for a certain amount of anticipation. This kind of method can be used in current big data acquisition to figure out what kinds of topics or things the user has been interested in in the past, and from that we can also infer what the user will be interested in. But Weibo information changes very quickly, and users can suddenly change with the heat. So we also need to make more comprehensive observations of the user's background or the user's characteristics and so on.

4.2. Forecast the interesting degree of reading the promoted content information

The reason why users forward this content is definitely because he is very interested in this content. We can analyze what kind of content users have forwarded in the past, so as to know which things the user would be more interested in. Of course, this kind of prediction also needs to be comprehensively explored through the age, social status, income and so on . In addition, many people these days seldom make substantive comments about content, so it is difficult to find out where the actual pain points are.

4.3. Through the prediction of group influence

The people you make friends with, the things you come into contact with in your life, or the topics you talk about, are framed to a certain extent. In order to have the same topic with your friends, you will have to read about topics that interest this group. It is more difficult to predict such impacts than the previous two, because we cannot enter in our personal lives, and it's harder to know what kind of people you're with. However, certain information can still be obtained through people who are concerned and those who are concerned.

4.4. Construct feature vector to predict formula

This paper will combine the characteristics of the research object through coding and calculation methods. Finally, the quantitative formula can be obtained. We can take the user's data, text, or other information, and put it into a certain form, which is the encoding form mentioned above. We can

specify in advance which values are the reference values, and on this basis, we can refer to formula (1). By studying the activity of users, we can probably know when and what content users have the highest frequency of occurrence, and can infer a certain reference value.

$$\text{user liveness} = \frac{\text{number of users' microblogs}}{\text{current time - user created time}} \tag{1}$$

The calculation is mostly about content. For content, a Bayesian classifier is used. We need to find out the key word and make sure it's the accurate one. This study uses the mature word segmentation system to carry on the segmentation. When the participle is finished, we get the word set of the content. Next, use a pre-prepared training set to work on the topic. But such a sample is not the text content that the actual user is interested in. However, it is not suitable for collecting data samples commonly used in machine learning, which requires the actual interest of users to collect data. It takes a lot of time and manpower to conduct accurate research.

Here we will calculate three types of similarity: publisher and recipient tag similarity, topic and recipient tag similarity, and topic and author tag similarity. Such a calculation can not only calculate the similarity between the two words, but also can be a certain conversion. We all know that web users usually have not one tag, but many. So when calculating similarity, suppose the set of tags SW1{ W11, W12, W13, . . . , W1n } for micro-bloggers, the set of tags SW2{ W21, W22, W23, . . . , W2M } for micro-bloggers The similarity between the tag of a micro-blogger and the tag of a micro-blogger is given by formula (2).

$$S = \frac{\sum_{i=0}^{n} \sum_{j=0}^{m} \text{Sim}\left(W_{1i}, \ W_{2j}\right)}{m * n} \tag{2}$$

References

[1] Data source: cnnic and China Commercial Industry Research Institute

[2] Gao S, Ma J, Chen ZM. 2015. Modeling and predicting retweeting dynamics onmicroblogging platforms. In: Cheng XQ, Li H, eds. Proc. of the WSDM .

[3] Zaman TR, Herbrich R, Gael JV, Stern D. Predicting informationspreading in Twitter. In: Proc. of the NIPS Workshop on Computational Social Science and the Wisdom of Crowds. 2010.

[4] Xu ZH, Zhang Y, Wu Y, Yang Q. 2012. Modeling user posting behavior onsocial media. In: Hersh W, ed. Proc. of the SIGIR 2012.

[5] Boyd D, Golder S, Lotan G. 2010. Tweet, tweet, retweet: Conversationalaspects of retweeting on Twitter. In: Proc. of the 3rd Hawaii Int'l Conf. on System Sciences (HICSS 2010).