Media Leaders Mining Based on Android News Apps

Sijia Wang, Miao Zhang

Cyberspace Security School, Beijing University of Posts and Telecommunications, Beijing 100876

Abstract: With the rapid development of the mobile Internet, the mobile news apps have become the most important way for the public to obtain news. As a new media carrier and communication platform,the mobile news apps can promote the rapid dissemination of information and the rapid spread of influence. Some media have a major influence on the direction of other media reports and the behavioral decisions of the public. These media can be regarded as media leaders. Media leaders are very important in the dissemination of news. By identifying media leaders, companies or governments can promote sales or guide public opinion separately. This article believes that media leaders mainly achieve their own influence by publishing news, so this article uses the news published by the mobile news apps as an entry point. This paper firstly solves the problem of data crawling in mobile news apps, and proposes a data crawling method based on reverse analysis, and obtains the data source. Then, reconstruct the reprinting path of the news, and carry out accurate traceability. Finally, cluster the news based on LDA, and propose an algorithm for mining media leaders from three aspects: influence, activity and preference. Experimental studies of data sets have shown that our algorithms can effectively identify media leaders.

Keywords: Media leaders; News apps; Reverse analysis; Traceability

Introduction

In the first half of 2017, the number of mobile phone users has reached 1.1 billion. The development of mobile phones is no longer just a communication terminal, but also includes many necessary functions in life and work. Mobile news clients have also become the first channel for us to obtain news reports and the fastest place for news dissemination. According to statistics, the number of news and information categories APP on the domestic Android platform has exceeded 600[1]. For mobile news client and web news content is different, there are three main points. First, many news APP do not have WEB sites; Second, even if the news APP has a WEB site, since mobile news client is the most popular way for the public to obtain news now, the developer will focus on the data update on the mobile side and take the mobile side as the exclusive publishing platform for news, the update of web news will be delayed or even missing, so the number of news content on the mobile news client will be more and more comprehensive. Third, for the same piece of news, if there is both a web page version and a mobile end, the number of comments and the amount of reading are not the same. Based on the situation, it is necessary to reconstruct the reprinting route of news first. By digging out media leaders, it can provide good support for relevant government departments to master information and implement efficient and scientific decisions.

1. Related research

In recent years, there has been relatively little research on the communication of mobile news clients both inside and outside the network, mainly for social networks. In the field of opinion leader mining, measurement is mainly based on social network topology, user behavior and interactive information. In the topology of social network, the graph model is mainly established according to the relationship between users, and then according to the different analysis

methods.

It can be divided into two directions: one is user-centric analysis based on complex networks, and the other is relative ranking method based on random walk. The analysis method based on complex network can analyze the node degree and the shortest path. The node degree

The quantity method includes the degree of entry, degree of exit and degree of centrality, while the shortest path measurement method includes the degree of compactness and the degree of betweenness centrality. Centol found that the higher the clustering coefficient of nodes in social networks, the greater the influence of nodes[2]; Kitsak uses K kernel decomposition to divide nodes into different levels. Nodes with Ks value of 3 are nodes[3] with great influence. The measures based on random walk include PageRank, HITS and LeaderRank. These methods regard the influence of users as random walk, and the influence of nodes is related to the influence of other nodes connected with them. Weng and others discovered that users are in social networks has the feature of homogeneity. According to the similarity of topics, the PageRank algorithm is improved, and the TwitterRank algorithm is proposed to calculate the user's influence[4]. However, these social network topologies are static, and all explicit social relationships from the beginning of the birth of social networks, that is, the social relationships established by users a few years ago and those just established, are treated equally. Moreover, in such a network topology, the weights of all connections are the same, and the influence among connected users is the same. This is inconsistent with the actual situation and does not take into account the user's interaction behavior, resulting in a deviation between the measurement results of the method and the actual situation.

In the aspect of user interaction behavior, the influence of users is mainly reflected by analyzing the influence of information sent by users and its propagation characteristics, thus mining opinion leaders. Agarwal and others comprehensively considered the number of references, comments, novelty and content length of blog posts to evaluate their influence[5]; Li and others comprehensively considered the quantity and quality of blog posts to dig out opinion leaders[6] in commodity evaluation. Fan Xinghua and others proposed an influence probability diffusion model[7] to measure the influence of forum posts by improving the influence diffusion model.

In the aspect of user interaction information, Wu Yu *et al.* calculated the user attribute value through iterative algorithm, and then calculated the users influence[8]. Ding Xuefeng *et al.* adopted the method of constructing the users attribute matrix and ranking the comprehensive weights[9]. Yang *et al.* considered the influence of non-adjacent nodes, and proposed the algorithm[10] to measure the users influence based on the intermediate relationship. Matumura *et al.* proposed the influence propagation model IDM[11] based on the similarity between the reply content and the message published by the user.

The above methods are mainly aimed at social networks. However, the way information is transmitted in mobile news clients is different from the way information is transmitted in these environments. At present, the research on media mining in mobile news clients is almost blank, resulting in insufficient research on information transmission in mobile news clients for the following reasons:

On the one hand, there are still big problems in news collection on mobile news clients. The API of the mobile news APP based on the C/S mode when communicating with the corresponding news server is not public and cannot be directly obtained. Therefore, when capturing news on the mobile news APP, the portal URL for capturing news cannot be directly specified. In addition, when the mobile news APP communicates with the news server, the request usually contains many parameter signatures. If the signature algorithm of these parameters is not known, the crawler will not be able to simulate the request to communicate with the news server and will not be able to capture the news content in the news APP.

On the other hand, there are social relations between users in social networks that are concerned and paid attention to, but there is no such social relation between mobile news clients. In addition, the dissemination of information in social networks is mainly realized by users' forwarding, while the news content in mobile news clients is mainly carried out through news reprinting by mobile news clients, while mobile news clients influence the public through the reprinted news. The public does not participate in the news dissemination process, and the dissemination node is mobile news clients. Therefore, the information dissemination mode in mobile news clients is obviously different from social networks. First, in the news dissemination of mobile news clients, the number of nodes is small, and there is no social

relationship, which cannot form a social network. The second is that the spread of news is mainly through the news client to reprint the news, users do not participate in this process, and the influence on users cannot be measured in the constructed forwarding behavior network.

2 Research on media leaders mining

2.1 Mobile news client classification

According to the content, mobile news clients are divided into the following categories:

(1) Traditional media news APP

This kind of mobile application mainly refers to the timely transformation of traditional media in order to adapt to the development wave of mobile Internet. The independently developed mobile news application is dependent on its own original media brand, mainly including local news media and central news media. The mobile news clients corresponding to local news media include: Beijing News, Xinmin Network, Ai Ji'nan, etc. These news information APP have strong geographical features and pay more attention to local news. The mobile news clients corresponding to the central news media include: People's Daily, Xinhua News Agency, China Youth Daily, Health Times, etc., which mainly report on news happening nationwide.

In addition, the traditional media group has also taken further reform measures, pulling out some personnel to do independent mobile internet news information APP to create a brand new news product, instead of relying on the original media brand. For example, interface news, cover news, etc. These mobile news clients rely on media units and have the right of independent interview. They have unique advantages in reporting and tracking current events.

The content push of traditional media news APP mainly includes the news content produced by its own media and the news content of other media reprinted in small quantities. This kind of news APP depends on the corresponding media to generate news content, including the central news unit, ministries website, provincial news unit, etc.

(2) Aggregate news APP

This category of news applications can be divided into two categories: one is news portal transplantation, which has taken the lead in the market of mobile news clients and launched fierce market competition on the mobile side, such as Tencent News, Netease News, Sohu News, Sina News, etc. The other is comprehensive news media, such as today's headlines and a little information. They mainly calculate and push the news that users are interested in through specific algorithms. They are the grabbers and integrators of massive news content.

In these mobile news APP, self-media, also called personal media, have been launched one after another. It focuses on popularization, individuation, electronization and self-dissemination. Everyone can create news, everyone is the media, breaking the limitation of time and region. Users can also become creators and disseminators of news. The entry threshold is low, making it a reality for the common people to establish their own media. These aggregated news APP rely on self-media to generate news content.

(3) Vertical news information category APP

For financial, health, science and technology, sports, entertainment and other vertical fields, provide a platform for some groups to collect and integrate news information. Its goal is different from the aggregate news APP. It does not pursue the quantity and comprehensiveness of news, but is specialized in specific fields, such as tiger-bashing sports and orange entertainment. This kind of news APP usually has its own editorial team to publish original content.

2.2 Data acquisition based on reverse analysis

Since the API of communication between mobile news client and server is not public, there are still relatively big problems in news collection on mobile news client. The crawler cannot simulate the request to communicate with the news server and cannot capture the news content in the news APP. Therefore, it is necessary to adopt the method based on reverse analysis to collect data.

Before collecting news information, you need to obtain the interface called when opening the news page. First, you need to use proxy to grab the API of communication between mobile news client and server. API is mainly divided into two categories, one is the API of the catalog page and the other is the API of the news specific content page. The API of a catalog page refers to a catalog page containing a large number of news specific content pages API; The API of the

specific news content page points to the actual news content page, in which we can grab information such as the news title, text, number of comments, author name, etc.

Since many parameters in the API are transmitted after encryption, it is necessary to obtain an encryption algorithm to simulate the communication between the client and the server. In this paper, static analysis and dynamic debugging are combined to obtain the encryption algorithm of parameters.

Android reverse analysis can be divided into two methods[12] according to the generated code after decompilation. The first method is to decompile the classes.dex file into Java code. The advantage of this method is that the converted Java code is clear and easy to understand, but it cannot be reassembled to generate APK and can only be used as an auxiliary code reading method „and this reverse method is easily confused, thus causing the generated code to be very chaotic and even unreadable code to exist. The second method is to decompile the classes.dex file into Smali code. Smali code is an object-oriented assembly code with strong readability. The modified Smali code can also be repackaged and signed to generate a new APK file. In this system, the second reverse method will also be used. The commonly used tool for reverse coding into Smali code is JEB. this tool set understands compression, Smali positive and negative compilation, assembly, etc.

2.3 Topic clustering based on lda topic model

Latent Dirichlet Allocation (LDA) occupies a very important position in the subject model field. It was jointly proposed by David Blei, Andrew Ng and Michael Jordan in 2003 to calculate the subject distribution[13] of text content. Since the LDA model was put forward, it has been very popular in the field of text processing and has been continuously developed. LDA model is in PLSA

It is improved on the basis of the Probability Latent Semantic Analysis model. In the PLSA model, the word distribution and the topic distribution are both definite and unique, and no probability model is provided in the text processing process. However, in the LDA model, the word distribution and topic distribution are both variable, based on Bayesian probability. Word distribution and topic distribution belong to polynomial distribution. Because polynomial distribution and Dirichlet distribution are conjugate, Dirichlet distribution is used in the conjugate prior distribution of word distribution and topic distribution in LDA model, and a text-level probability model is further proposed.

The core of LDA model is a three-level Bayesian model, on which the probability distribution of words, topics and text content is described. By creating potential topic layers between documents to model documents, potential topics in the corpus can be discovered. It is a Bayesian network that generates documents using various topics, gives the topics of each text content in the text set in probability, and also calculates the important words in each topic. All potential topics are not of equal probability, and some topics may have higher absolute probability values than others. If the score of topic importance is higher, it means that the topic covers the important part of the article[14].

2.4 Sources of news

In the information dissemination of mobile news clients, each news client will have its own corresponding crawler, which will continuously capture the news data of other news clients. It is usually aimed at a topic. Some media report it first, while other media concerned with the topic reprint the report selectively, and the news spreads. If the media publishing original news is regarded as a root node, each media reprinting news is regarded as a child node, and each reprinting behavior is regarded as an edge connecting different media, then the reprinting relationship of a piece of news can form a reprinting tree.

According to the captured news data, the reprint relation is reconstructed based on the explicit annotation reprint relation[15]. Annotation-based data dissemination analysis method refers to metadata dissemination after adding annotation information in the dissemination process. Through analysis and mining of the data in the dissemination process, the annotation information can be found and the current data source can be found. Due to copyright issues, news should be marked with the source of reprint when reprinting, and the reprint route should be reconstructed according to the reprint information so as to trace the source of news.

Under a topic, there will be many different original news, so the transmission tree of a topic is made up of several original news transfer relation trees. First of all, the list of all original news about the topic is obtained after the topic clustering. Then the reprint tree of each original news is reconstructed in turn. Finally, the information dissemination

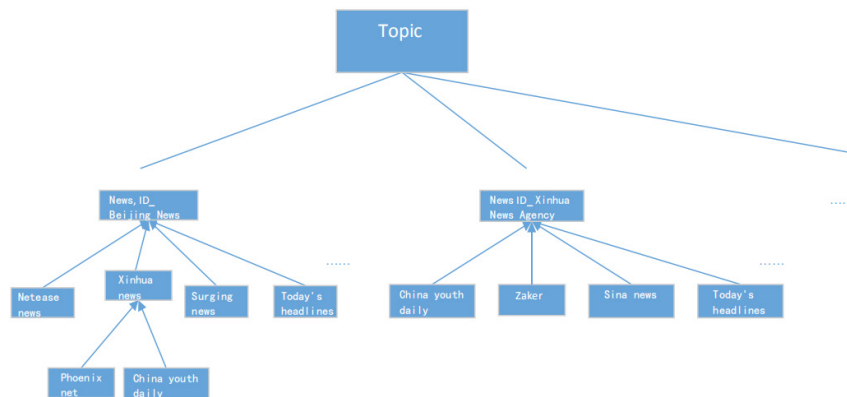tree of a topic is reconstructed, as shown in Figure 1.

Figure 1. Information dissemination tree for a topic.

2.5 Analysis of media leaders

In the news dissemination of the mobile news client, the news is reproduced by the mobile news client and then read by the user, which affects the user. The diffusion of information is mainly formed by media reproduction. In a huge amount of news, some news is reprinted by many media, causing many users to read and comment, and even having a great impact on the topic orientation within a period of time. However, some news is not paid attention to by users and other media. In addition, in different topics, influential news publishers are different, which shows that the influence of news media is sensitive to topics. In this article, it is believed that the influence of mobile news client is mainly displayed through original news. If the news client A publishes an original news and is reprinted by the news client B, this article counts the influence of this news on B as A because the root of the influence is the original news itself.

Generally speaking, influential media have some typical characteristics. This paper proposes a media leader mining algorithm to mine leaders in mobile news clients by analyzing influence, activity and preference.

2.5.1 Influence

The influence of news APP is mainly reflected by the influence of original news released, and the two main factors that determine the degree of influence of a news are media attention and user attention. Generally speaking, the research on news influence mainly focuses on these two factors. Therefore, the measurement of news influence can be divided into three categories: news influence calculation based on media attention, news influence calculation based on user attention, and news influence calculation based on media attention and user attention. This paper uses the third method to calculate the influence of news. In addition, timing can also reflect the media's sensitivity to news.

(1) The media attention will cause other platforms to enter when news is released on one platform in an era when the subsequent media reprinting mode is prevalent.

Line reprint or tracking reports. The number of times the news is reprinted is an important criterion to measure its influence on the media. If the number of times the news is reprinted to other mobile news clients after the news release, the stronger the influence on the media, the more users will be indirectly affected and the influence will be expanded. In this article, we use the number of times news is reprinted to express media attention. According to the information dissemination tree of a specific topic, reprint information of news can be obtained. In this article, mediainf (c) and i refer to the number of times news c and i have been reprinted.

(2) User attention

The number of times the news has been reprinted can only indicate the extent to which the news has spread and how many mobile news clients the news has been published. The number of reprints describes the spread of news, and can only measure the influence of news in breadth, but not completely. The purpose of publishing news is not only that the news is reprinted by other media, but also that the news has influenced it.

How many users, news depth influence is reflected in the influence on users. Comments are generally located at the bottom of the news and are displayed in the form of waterfalls. Users post comments after the news theory, which is carried out spontaneously by users, has strong individuality and randomness, and belongs to a popular expression of opinions. This formula can make users express their views on news conveniently and enhance the interaction between

news content and users. Therefore, the number of comments can reflect the attention of users to a certain news. The more comments, the greater the participation of users, and the greater the influence of news on users. In this article, the number of comments is expressed as userinf (c) and I as indicators to measure the impact of news on users.

(3) Each news item with time sequence will have its own release time. In theory, the earlier the news item is released, the more likely it is to have a greater impact. Therefore, timing is a very important indicator in the judgment process. According to the release time of all relevant news about an event, this article is arranged in descending order, and the later the release time, the lower the ranking.

Before, the earlier the release time, the lower the ranking. Time (C) and I indicate the temporal characteristics of news C and I. The larger the temporal attribute characteristic value of news, it shows that in the same topic, the news reported a new event earlier.

2.5.2 Activity

Activity reflects the media's initiative to participate in the process of topic dissemination, mainly through the publication of original news about specific events to express their views and opinions, as well as through reprinting news to spread information.

(1) Number of original news

The number of original news is used to measure the willingness of news APP to express opinions and opinions voluntarily. The number of original news published on a specific topic reflects to a large extent the activity of the news APP on the topic. The indicator is represented by num (app) and i and app and i represent the current mobile news client.

(2) Number of news reprinted

News, APP and news reprinted on a specific topic reflect the communication power of the mobile news client on this topic. To a certain extent, they can reflect the activity of the news APP and APP on this topic. The indicator is represented by Reprinnum (APP) and I), and APP and I represent the current mobile news client.

2.5.3 Activity

Under a specific topic, in order to calculate the media's preference, the media's preference to this topic is measured by calculating the proportion of the number of new news included in the specific topic in the media to the total number of news in the media, which is expressed by Numper (App I). To some extent, the higher the proportion, the stronger the media's preference for a topic.

2.5.4 Weight determination

The weight coefficient of each attribute is another factor that affects the accuracy. In this paper, AHP multi-criteria decision-making method combining qualitative and quantitative methods is used to determine[16]. Firstly, according to the criteria of equally important, slightly important, obviously important, much more important, extremely important and so on, the attribute indicators in each level are compared in two ways, and the 1-9 scale method is adopted to determine the relative importance of each indicator in each level and obtain the judgment matrix. Secondly, consistency check is carried out. Then, the eigenvector of the judgment matrix is calculated by the square root method, and the weight coefficients of attribute indexes in each level are obtained, and then the weight coefficients of influence, activity and preference 3 levels are calculated respectively.

3 Experimental analysis

3.1 Data collection

JEB is a commonly used decompilation tool. Based on interactive visual operation, it carries out full-automatic decompilation. It has strong forward and reverse indexes and a certain degree of renaming capability and searching capability[17]. In this system, JEB is used to decompile apk. By using proxy to grab communication API, as shown in figure 2, it is packets of different categories of entries, where sign, time and menuid are variable parameters and need to be constructed.

Figure 2. Communication API.

Because the parameters in the above request will be relatively close to each other in the source code, the parameters will be searched in the decompiled source code thumbrate, and then find the construction method of sign parameter, as shown in Figure 3.

```
private Params createParams(Params arg8) {
    if(arg8 == null) {
        arg8 = new Params();
    }

    arg8.put("clientid", "1");
    arg8.put("ip", NetworkUtil.getLocalIP());
    long v0 = System.currentTimeMillis();
    arg8.put("siteid", this.ctMediaCloudConfig.siteId);
    arg8.put("system_name", "android");
    arg8.put("type", "android");
    arg8.put("sign", SignUtil.getSignStr(arg8.getUrlParams(), this.ctMediaCloudConfig.secretSign, v0 + ""));
    arg8.put("time", v0 + "");
    return arg8;
}
```

Figure 3. Static debugging.

Based on the above static analysis, a breakpoint is set at the getSignStr function in the source code for debugging, and the string 2c0dbc39b6e9dd710dbf8952291b0a7eda9706d0faccbf382797d5c16240662015OO625743840 is encrypted with MD5. 2c0dbc39b6e9dd710dbf8952291b0a7e is the value after MD5 encryption of the result string, da9706d0faccbf382797d5c162406620 is a secretKey and is the constant parameter defined in the source code, 15OO625743840 is the timestamp, as shown in Figure 4.
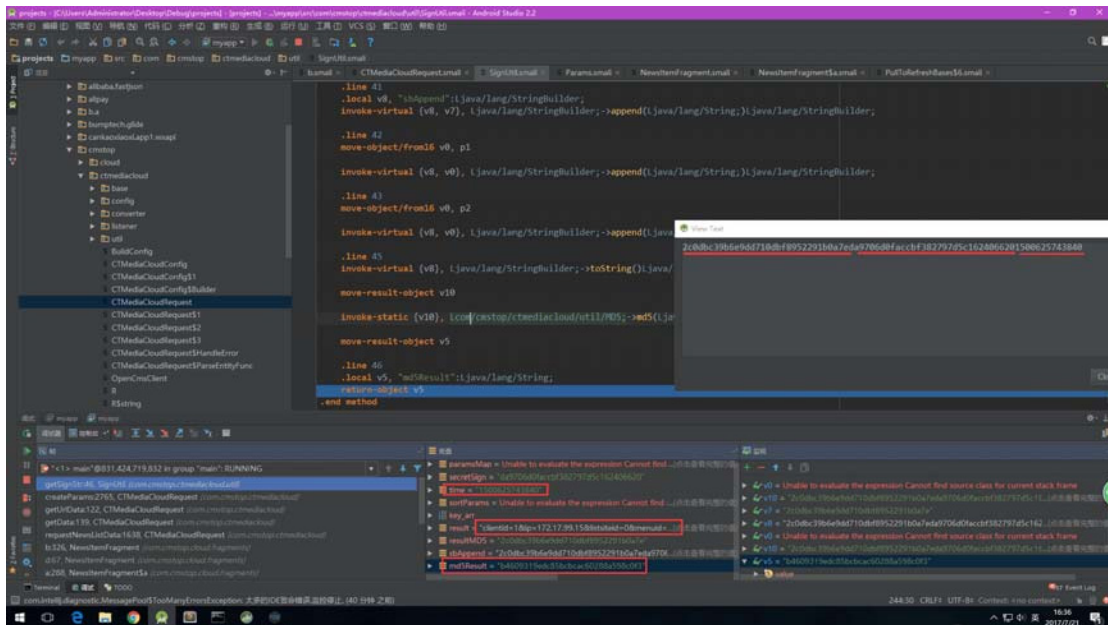

Figure 4. Dynamic debugging.

3.2 Analysis of experimental results

3.2.1 Data set

Based on the reverse analysis technology, this paper collects the news data from mobile news clients and clusters topics based on the LDA main topic model. This paper takes Chongqing bus falling into the river as an example to analyze the relevant broadcast data of the event in a certain time range and mine media leaders under specific topics. Among them, each topic includes all the news information within the topic, and each news information includes detailed information such as news title, content, author, originality, reprint source, number of comments, etc.

3.2.2 Traceability analysis

In order to facilitate the research, the original news with less than 5 reprints and less than 50 comments has been removed. By tracing back to the source of the news articles and reconstructing the reprinting route, the research found that for serious news, the revision of the title in the reprinting process tends to be positive.

In the process of news reprinting, the staff of the mobile news client will generally modify the title, and it is worth exploring whether the problem of marking party has been aggravated in the process of modification.

This article mainly focuses on social, domestic and technological news. These news are all serious news, and the original authors of these news are basically traditional media, and then reprinted by other mobile news clients. The original titles are marked by traditional media. In the process of reprinting, the titles are modified through online editing. In this article, taking today's headline as an example, the matching degree between headline and news content is calculated to measure whether the phenomenon of marking party is aggravated. In the public's impression, the new media are more inclined to change the title to an eye-catching style, regardless of authenticity and users.

Experience, however, such conclusions are more likely to be applied to non-serious news. For serious news, the new media will instead pay more attention to the consistency of the title and news content, because the title of the news to a certain extent determines whether users will click on the news, so the information contained in the title will be more comprehensive.

3.2.3 Media leaders mining analysis

In the following part, we will make an in-depth analysis of the situation in the statistical results.

(1) Compared with mining media leaders only according to a single index, the algorithm in this paper is more reasonable. As can be seen from the table, Xinhua News Agency has a higher ranking in terms of activity and preference. However, due to published articles chapter influence is relatively small, so the comprehensive ranking calculated by the algorithm in this paper is relatively low. After confirmation, it was found that although Xinhua News Agency has high activity and preference, it has covered a total of 318 news articles in response to the Chongqing bus crash incident, which is the highest number of news in all news clients and focuses on the incident. However, the vast majority of these news articles are reprinted and only a few are original. Xinhua News Agency did not send reporters to follow the development process of the incident, so it did not release many influential news articles, mostly follow-up reports and articles summarized after a few incidents, which have both influence on users and influence on the media.

Judging from people's past experience, it may be thought that Xinhua News Agency will play the role of media leader in this incident, but through scientific calculation and reasonable analysis in this article, Xinhua News Agency's ranking in this incident is not high.

In addition, the news released by China Youth Daily has great influence, but it did not actively participate in the dissemination of the event, nor did it focus on the development process of the event. Its activity and preference are relatively low, so its comprehensive ranking is relatively low.

The influence and activity of today's headlines are relatively high, but the preference ranking is relatively low, which also reflects the characteristics of the customer end of aggregated news and will not focus on following up the development of the event. In the overall news, only a small portion of traffic can be divided to report the event, and there is no focus. However, even so, the overall ranking is still high due to the high quality of news released by today's headline client and the active dissemination and diffusion of news related to the event.

The above analysis results show that when judging media leaders, it is not reasonable to only consider the influence of news, but also the attributes of media itself, which can be reflected by liveliness and preference. The influence pays more attention to the influence of original news released by mobile news clients, the activity pays more attention to the active participation of mobile news clients in reporting the event and the degree of spread, and the preference pays more attention to the authority of mobile news clients in the event field. Only by calculating these three characteristics can media leaders be comprehensively evaluated.

(2) In social news, traditional media have greater news influence and self-media have greater activity. In the news dissemination of mobile news clients, the central media and local media have not declined, but are active in the competition.

Reform and release more standardized and higher-quality content. Even in aggregated news clients, the most influential news is the news released by reprinted central media and local media. The influence of original news from media is relatively much lower. Original news released by traditional media has relatively greater influence on users and media.

4. Conclusion

In the era of mobile Internet, mobile news client has become the most important and fastest channel for the public to obtain news. It is also a place to pursue timely news and an important way to distribute news content. It is very important to excavate media leaders for news. It is helpful for the government or celebrities to have a clear understanding of the news dissemination process. It can guide public opinion and guide the direction of news reporting through news media during the dissemination process. In this paper, the whole mobile news client ecosystem is traced and analyzed, and then media leaders under specific topics are mined, and analyzed on real data sets.

Follow-up work can consider mining implicit reprint when constructing the news reprint path, because the news market is not standardized enough, some news reprints will not be labeled with the source, or will be published as original articles after modification, which makes the traceability process not accurate enough, and digging out this situation will make the calculation results of media influence more accurate.

References

1. Media Consulting ,.2017Q3 ,Quarterly Monitoring Report of China Mobile App Store, [OL]. [2017-11-23]. http://www.iimedia.cn/59804.html

2. Centola D. The spread of behavior in an online social network experiment[J]. Science, 2010, 329(5996):1194-1197.

3. Kitsak M, Gallos L K, Havlin S, *et al.* Identification of influential spreaders in complex networks[J]. Nature Physics, 2010, 6(11):888 -893.

4. Weng J, Lim E P, Jiang J, *et al.* TwitterRank:finding topic-sensitive influential twitterers[J]. 2010.

5. Agarwal N, Liu H, Tang L, *et al.* Identifying the influential bloggers in a community[J]. 2008.

6. Li F, Du T C. Who is talking? An ontology-based opinion leader identification framework for word-of-mouth marketing in online social blogs[J]. Decision Support Systems, 2011, 51(1):190-197.

7. Fan Xinghua ,Zhao Jing ,Fang Binxing ,*et al.* Probability Model of Diffusion of Influence and Its Application to Research on Discovery of Opinion Leaders [J]. chinese journal of computers ,2013, 36(2).

8. Wu Yu, ma lulu, Lin Mao, *et al.* opinion leader discovery algorithm based on user influence [J]. minicomputer system, 2015, 36(3):561-565.

9. Ding Xuefeng, Hu Yong, Zhao Wen, *et al.* Research on Characteristics of Internet Opinion Leaders [J]. Engineering Science and Technology, 2010, 42(2):145-149.

10. Li Y, Qiao Y, Liu Z, *et al.* Identifying opinion leader nodes in online social networks with a new closeness evaluation algorithm[J]. Soft Computing, 2018, 22(2):453-464.

11. Matsumura N, Ohsawa Y, Ishizuka M. Influence Diffusion Model in Text-Based Communication[J]. Transactions of the Japanese Society for Artificial Intelligence, 2002, 17(3):259-267.

12. Li Tianhui, Wang Yanqing, Android, Reverse Analysis of Malicious Code, J. Information Security and Communication Secrecy, 2016(4):95-98. "

13. Blei D M, Jordan M I, Ng A Y. Hierarchical Bayesian Models for Applications in Information retrieval[J]. Bayesian Statistics, 2003:25-43.

14. Zou Xiaohui, Sun Jing. LDA Thematic Model [J]. Intelligent Computer and Applications, 2014, 4(5):105-106.

15. Minghua, Zhang Yong, Fu Xiaohui. Summary of data traceability technology [J]. Minicomputer System, 2012, 33(9):1917-1923.

16. Lu yuejin. Ranking of fuzzy analytic hierarchy process based on fuzzy consistent matrix [J]. fuzzy systems and mathematics, 2016, 16(02):79-85.

17. Chen Peng, Zhao Rongcai, Single Sign, Etc. Android Based on Combination of Static and Dynamic, Malicious Code Behavior Similarity Detection, J. Computer Application Research, 2018, 35(5).