An SNS Mechanism to Highlight Low-profile Information

TAISHI NEMOTO¹, TAKAYUKI FUJIMOTO²

¹ Graduate School of Information Sciences and Arts, Toyo University, Tokyo, Japan
² Graduate School of Information Sciences and Arts, Toyo University, Tokyo, Japan

Abstract - In recent years, with the spread of smartphones and the Internet, users have been able to freely leave various forms of information on the Internet. In particular, social media is full of information posted by many users. The flood of information can mislead them. You will see, for example, different users sharing posts that are not really different from one another. They are about something that many people show interest in or talk about. Or they can be about the news of different areas, such as entertainment, politics or economy. Content curation services/websites are also popular. They repost and re-edit content that made a huge impact. A famous one is Togettr which groups together Tweets on the same topic. There is also a curation site that picks and displays the posts of the 2channel text board. On the Internet, you can “reprocess” the content by organizing the information in an easy-to-see, interesting way. Posts on social media, or SNS are arranged in chronological order and shown in timeline, but the only way to put them together by content or topic is to manually search for keywords. If posts can be classified and put together by content, SNS would be useful in gathering information, not only as a communication tool. In addition, some efforts would be needed to make low-profile posts as visible as influential posts. Niche, less viral content tends to be forgotten, but biased display of content is not desirable for those who gather information. We need a chance to be exposed to unexpected information. In this paper, by putting together the content related to influential keywords, we propose a social networking mechanism that displays posts, giving priority to those with low impact and slim chance of being viewed, but also with originality.

Keywords: Social media, Information gathering, Smartphone, Messenger, Communication

1 Introduction

The number of social media users is rapidly increasing worldwide. Most of them are users of major social media. For example, the number of monthly active users of Facebook is 6.5 billion, which is remarkable. In addition to text-based media, photo-based and video-streaming applications are also emerging. However, their format is not really different from the existing ones. Most of the social media are time-based, and most are arranged in chronological order called timeline. However, for users who also value old information, the information from the media is often very important. New information tends to stand out in SNS where old information is buried. This characteristic of social media is a critical defect in terms of information gathering. In this paper, we aim to devise a system with which users can reprocess and organize old information. In recent years, # (hashtag) has become common as a way of organizing information. Having been used for a long time to categorize information, hashtags make it more convenient for people to search and view the content. However, keywords in such tags are not always helpful in topic-based classification. A large number of keywords are used in posting on SNS, and classifying by each word does not work very much. Looking at the context, there are many cases where the same keywords are used or the words associated with the keywords are slightly off the point.

2 Background

The strength of social media is its immediacy of capturing from many users the trend and a large amount of information. With the spread of smartphones, the use of SNS to collect information has now become commonplace. It is said, however, that there are some problems with SNS. It is called cyber cascade problem or echo chamber phenomenon. This is a phenomenon typically seen in social media, which strongly reflect the characteristics of the Internet. Although the advantages of SNS are its large amount of information and immediacy, if the information is quite common and not unexpected, it seems that, when gathering information, the novelty and efficiency of SNS would be considered extremely low.

Unexpected information is important in information gathering. With only things known, it would not be much of information gathering. You just end up browsing what you want. In order for you to gather information more efficiently, it is important to take a heed to low-profile information that only a handful of people might watch, not just the trending items.

The mainstream of current SNS design is the timeline. Timeline is not a term specific to any of the social media, but the way information is displayed chronologically. The latest information is designed to be displayed in the upper part of the timeline. Older information, on the other hand, either takes a search to view it or could not be retrieved at all. It may be because there are too many posts with the same content or personal content, not because the older information is not worth viewing. It might turn out a more valuable design if, by putting together the trending content and the same topics, you could see other content than these.
3 Purpose

The purpose of this research is to put together the same content and related keywords, in order to highlight information that is not trending. The current mainstream structure is the timeline-like design which has users keep uploading the latest information. On the other hand, there should be information of great importance which is not necessarily so timely. Trending content does not always help in gathering information. Also, viewing a large amount of information which you may not need can cause a phenomenon called SNS fatigue. There is an urgent need to devise a system that brings together all the information gaining popularity in the society, and highlights niche information. That way, information collection would be more efficient, and users could obtain information other than known content.

It is certainly possible, as some people say, to view old information with existing systems, but it would be through websites or e-books rather than social media. More generations are now using SNS rather than WEB browsers in order to search information. SNS is in a better position to keep track of fashion and trends, and to promptly receive the latest news such as traffic jam information. This is also the reason why it is called the newest media. In this paper, we propose the mechanism of a system to efficiently collect information on social media. Specifically, using text mining algorithms, posts with similar content are consolidated, and other pieces of information are displayed in another form. In this way, re-editing the fragments of information from SNS makes it easier for users to view them.

4 Analyze

With the spread of the Internet, a word called long tail has emerged. It is a marketing concept used in the Internet sales. The traditional idea of Pareto principle (80/20 rule) was that 80% of the sales are generated by 20% of popular products and good customers. According to the long tail concept, however, the total sales of low-demand products (tail) is greater than the sales of top sellers. This idea is applied to the number of views on the web, which can be considered a long tail SEO in web search. The following is an image for reference.

In this case, the idea is that the total number of searches for small keywords is greater than the number of searches for what are called big keywords. This research attempts to put together big keywords on SNS, and highlight small keywords. It also leads to an idea of making the main theme more visible by putting together big keywords in order to highlight small keywords.

The Internet bulletin board is an example of the idea of putting together the content by thread. On the bulletin board, each thread has a fixed theme and title with which the content is basically aligned. Messenger and group chat develop from there. These focus on contacts, not on themes. Then it focuses on the content for the general public, such as those on Twitter. An illustration is shown below.

5 Precedent System

As an example of a system which classifies content by theme or topic, Slack is gaining popularity recently. Slack is a communication tool that is said to be useful for organizing information. It is because you can choose the content and set theme for each thread. However, deciding on the theme first and storing information in the thread is no different from existing bulletin board systems. Although it is useful as a tool specialized in social media, it may not be particularly novel. The goal of this research is a system that organizes information from the content that users posted without any specific purposes for the general public to view. The information posted by the users without any purposes can be harder to view than those posted for each topic. For example, when political and economic news are shown as hits, the social media should be flooded with the same content. Users would get stressed out gathering information if they had to keep seeing the same keywords. We would also need to change the way content is presented by changing the font size, color, style and orientation to reduce users’ stress.

Togetter is a Japanese website that rework the Twitter content into the form of a thread like a bulletin board. There is also Chirpstory, a website in English language. It is useful in
collecting information because users can grasp the content of Twitter's fragmentary information by theme. These media fall in between the bulletin board and SNS, and are of value for their ability to re-edit the information that individual users want to post.

Word cloud is another presentation method of visualizing text data in which keywords are extracted and processed. This is a method of selecting words with high frequency of appearance and of displaying them in a size according to the frequency. Its visualization that adds convenience to information gathering is commendable. The problem is that it also extracts meaningless words such as ‘and’ or ‘the’. Therefore, text mining involves the process of deleting data such as symbols called cleansing and the part of speech called ‘joshi’.

6 Mechanism

In the text mining method, it is possible to extract frequently-used words and inverse frequent words in conversations, as done in the tf-idf method, for example. More specifically, we would need to classify words and phrases by weighting each word, and related keywords can be grouped. The method is applied to this research. We will classify related words, and have the infrequent content displayed in higher positions to cause an unexpected element in information collection.

7 Proposal

In addition, this paper introduces a topic line different from existing timelines. We will process information by introducing a new presentation method. The current display format on social media is chronological. It is because SNS is where up-to-date information is posted in a chaotic manner. However, those posts can be fragmentary information, and are nothing but personal opinions or statements. Therefore, if the posts are reorganized by content and re-edited, they would be useful in gathering information. In this paper, we propose a topic line as a form of displaying information which has been organized by content. The following is its illustration.

Fig3. Existing timeline (left) and newly proposed topic line (right)

The timeline just displays posts in chronological order, which is in a piecemeal fashion. In order to collect specific information, the only way we could do is to use a filtering method and narrow down our search for a theme with words or tags. By using the topic line proposed in this research, a user can view the already compiled information.

8 Conclusions

This paper proposed a mechanism to weight information on SNS, and display the classified content. SNS has already spread as a tool for collecting information, and the number of its users is approaching the entire population. However, the information is still fragmented and chaotic. That is why curation sites for re-editing the information are emerging. If the primary information can be classified instantly, it would improve visibility as well as the efficiency of information gathering. Our future task is to implement the system using Twitter API, and carry out an experiment to measure the effect on information collection.

9 References


