Content Based Solution for Facebook Overload

MADHUR KUKRETI
NIKHIL AGARWAL
WEI SUN

1. INTRODUCTION

Social media has overrun the world. Facebook connects your friends, LinkedIn networks your jobs and Twitter embodies your insightful and mundane thoughts (for better or worse). Slashdot, Digg and Reddit have made news social, while traditional news sites have embraced services like Disqus to bring discussion to their articles. Quora and StackExchange crowdsource answers for your deepest questions and Amazon has made a commerce community through product reviews. Most news comes in the form of chronological feeds. Newspapers, magazines and journals release current news on a regular basis. Television and radio broadcast the news live as it occurs. Online services like Reddit and Twitter list trending topics, and Facebook feeds display your friends' news in real time. But the 24-hour news cycle guarantees that no topic, no matter how important, survives any of these institutions beyond a few weeks. We need to compartmentalize information through feeds because there is just too much information to process at any given time. As the pace of news has accelerated, so has the rate at which people consume news. News feeds filter out the noise, consolidating information the world considers important. Yet, according to a Pew poll, not a single news audience is well-informed about current events. Feeds do not prevent information overflow; in many cases, they exacerbate it. We need new paradigms of consuming and discussing the news, not because we need to understand everything that's going on, but because accepting that we can't comprehend everything allows us to focus on what we do know. But there's a massive elephant in the room. Facebook, in its attempt to make the world more open and connected, seems to instead make people more unhappy and less satisfied. Twitter is a lightning rod for criticism, a land of profuse and meaningless tweets. Even marketers question if social media is just a huge waste of time in the face of enormous valuations of social media behemoths. The average daily information dumped on a social network user is 54,000 words [2], equivalent to the length of the average novel. That textual content is only 63 percent of the information dumped on every social network user on a daily basis. Social Media overload is real, and it is a problem to health and productivity when abused. It started with RSS streams boosting the popularity of blogging by providing a uniform way to connect bloggers with their readers. Later, social networking sites adopted the stream model for the information shared by their users. Facebook has registered its users friends relationships building a social graph concept, Twitter has introduced nonreciprocal following relationship thus building global interest graph. Users of social networking sites continuously produce a stream of shared status updates, photos and links and receive a converging stream from all their friends or followers. The role of a social graph is to define the routes for the streams of information shared by the users. However, today a lot of active social media users complain that their streams have become too overloaded and hard to extract useful information. There are several reasons for this overload. Besides the general increase of usage and amounts of data shared every day on social networks, we distinguish the following two reasons: the growth of the number of connections in the social graph and automatic updates coming from applications. Facebook's social graph started originally as a tool to keep in touch with friends, has now become too unspecific and dense. Many Facebook users have more than 200 friends ranging from their real friends and relatives to people they have met at professional events and feeds from favorite brands and news sites. According to the

recent study of Facebook social graph, the average distance between vertices of its giant component was found to be 4.7. It indicates that individuals on Facebook have potentially tremendous reach. Shared content only needs to advance a few steps across Facebook's social network to reach a substantial fraction of the world's population. This growth of connections density is a natural process, because as soon as the user discovers another interesting source of information, they like to subscribe to keep up with it. But all those streams put together into a common timeline represent a huge, messy, hard to consume stream, while, in its depths, it still contains relevant information. In late 2011 Facebook introduced so called frictionless sharing for applications. Now applications can automatically post verb-based updates on behalf of its users about what users are doing within the application, without the user having to manually push a Share button. So that applications are able to post on users' walls as soon as they, for example, listen to a song on jukebox, play a tune on sound cloud, buy tickets on Everbrite, or go for a run using Nike+. This move by Facebook provoked a lot of criticism: obviously, such amounts of data pollute an already overloaded stream. Moreover, it raises privacy issues. We proposed a solution which reorganizes the layout of posts and news on Facebook. This will compartmentalize the user's stream into different sections e.g. news feeds, posts by friends and posts from liked pages. Under these sections, the order of posts or feeds will vary for every user. This variation or order of the posts or feeds will be decided by employing machine learning. Machine learning will include computing a prioritization score for every post taking the input from media stream for a particular user. These inputs will include different parameters along with their weightage values. This approach will help in segregating the unwanted stream from the desired one for every user.

2. SOLUTION DESIGN

As discussed in the earlier sections, Facebook's news feeds can be over-populated and confusing to the end user. The news feed of an average user having a couple of hundred friends and subscription to about twenty pages will be fairly informational. The aim of the proposed system is to provide the users with a very easy to use and well defined interface so that that they see whatever they want to on their home page. First we will discuss the basic model of typical Facebook profile. When users log into their account they see what is called the user's "Home" page. On this page, they can view posts from a) their "Friends" – friendships are initiated by sending a "friend request" to another user on Facebook and completed once that user accepts the request b) "Pages" they like – "Page" is a public profile created for business, organizations, celebrities etc. and c) from people they "Follow" – By following a user on Facebook one can receive posts from that user on news feed without having to add them as a friend. In sum, all these sources contribute towards the news feed. Facebook also provides an option to sort this news feed in two ways – 1) Top stories i.e. According to popularity and 2) Most Recent. However, these sorting algorithms work on all the posts simultaneously. This means even though the user can sort the posts that appear on his timeline, he receives a mixture of news from all the aforementioned sources. In fact this is one of the major causes of social media overload – too much information from different sources making it impossible for the user to assimilate anything in its entirety. Also, people often find themselves scrolling down to get feed from a particular page or a friend they like. Hence, displaying the news feed in a more structured way would be a step in tackling the problem of social media overload. On May 29, 2013 Google introduced tabs in Gmail. The Gmail "Tabs" segregated the inbox into five optional tabs - primary, social, promotion, updates and forums.

Fig. 1. Tabs in Gmail inbox

Under the "Primary" tab user can see messages from friends and family, "Social" tab contains messages from social networks, media-sharing sites, online dating services and other social websites, "Promotional" tab encapsulates promotional emails, deals and offers, "Updates" tab contains bill receipts, confirmations and statements and the "Forum" tab has all the messages from online groups, discussion boards and mailing list. This prevents important "Primary" emails to be piled on by other less important messages. We propose to apply a similar methodology to the Facebook news feed. Our aim is to develop a user interface wherein the posts are categorized into three categories namely – posts from friends, posts from pages and posts from news channels. After classifying the posts, they are sorted in accordance to user preference. So a few questions arise - how will the posts be segregated? How will they posts be reorganized under each category? What factors will be used to estimate user preference? The following 2 diagrams provide insight to these questions:

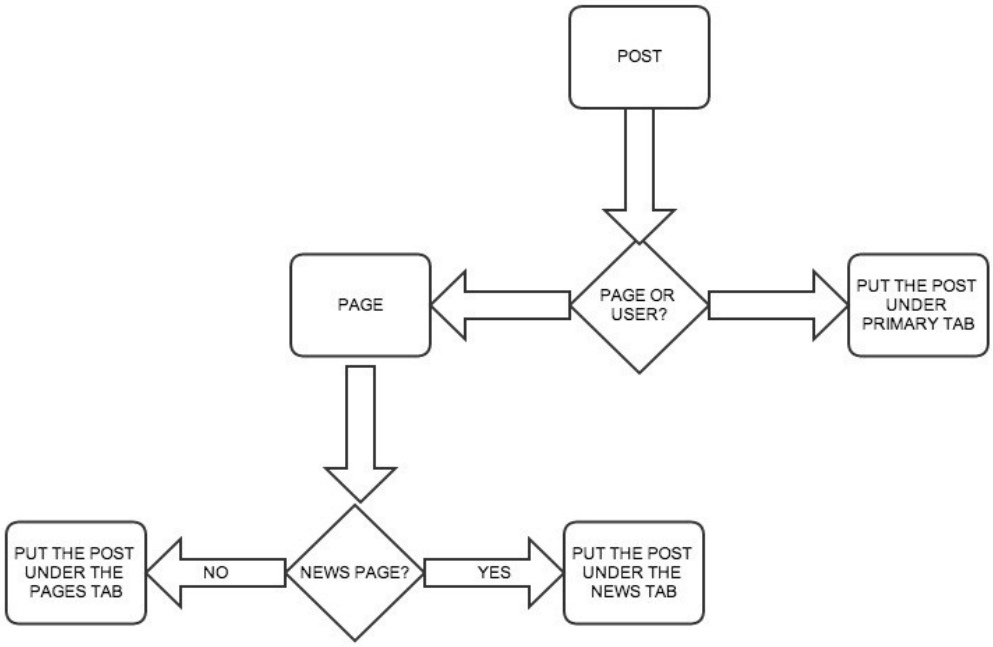

Fig. 2. Classifying the post

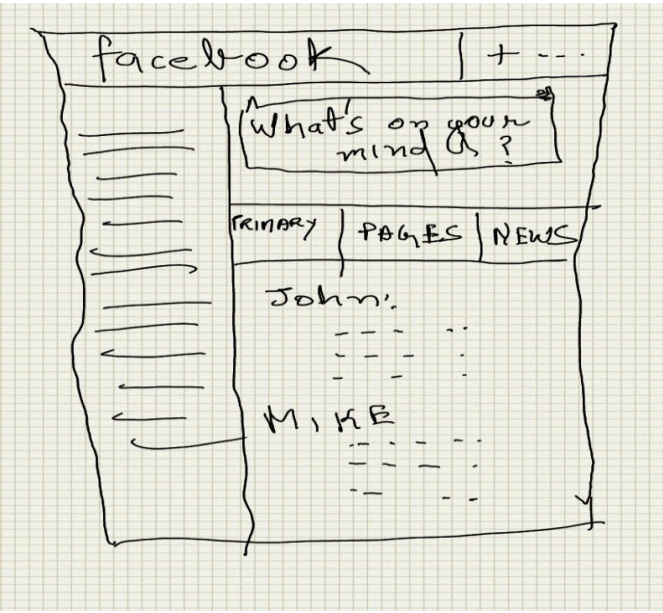

Fig. 3. Dummy UI after classification

Each post would assign one of the three labels – Primary, Page, News. This can be done in a few ways. First, Facebook can make it compulsory for every page, group or user to associate their post with either of these tags. Here Facebook can distinguish between a "User" and a "Page" implicitly. However, some of these pages can be news channel feeds. So to filter these pages out, Facebook can ask every page to declare itself as a news channel or otherwise. Another way to do this is using a local client on user's machine that can learn what kind of pages are news channels. This learning process will take 2 inputs – First, users can explicitly categorize and point out the news channels to the agent and second, the agent can try to identify news channels through the various attributes of their posts – For example, most the news channels have updates starting with – "Breaking now" or "Just in".

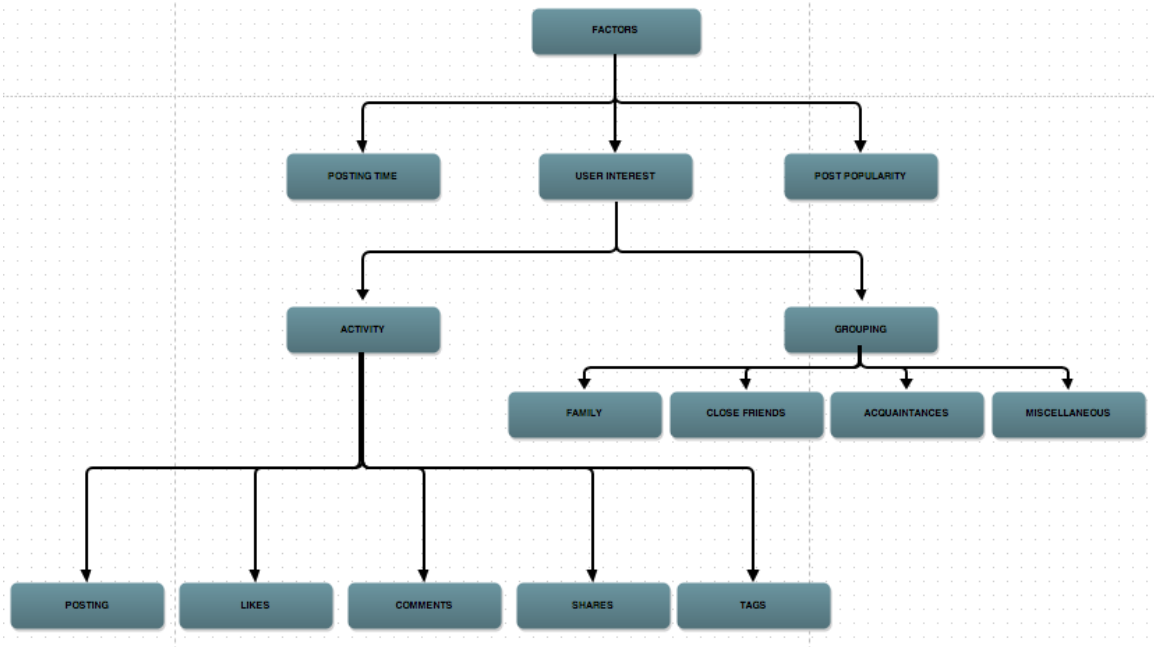

Fig. 4. Factors affecting the priority of the post

Next, the system will reorganize the posts under each tab according to user preference. How will this work? So in order to reorganize the posts, it is important to determine all the possible attributes of a post that give the closest approximation of user's liking towards the post. Some of the most common criteria are listed Fig. 4.

The priority of a post relevant to every user is judged by broadly three criterias. First criteria is user interest. Users prefer to see posts from people who are close to them and people who they like. Also, they tend to be more socially active with these people. On social networking sites such as Facebook, people tend to like, share and comment on posts of close family, friends and famous people. One way to judge this proximity between users is by dividing them into groups. In this case we consider 4 groups namely family, close friends, acquintaince and miscelleneous. Users explicitly provide this classification to Facebook. In addition, user's interest in a post can also be assesed by monitoring activity trends to determine what people are of interest to them (see Fig. 4). This factor along with "posting time" and "post popularity" are used to calculate the value of the priority number F(p) which is computed as follows:

$$F(p) = \sum w_i F_i$$

wi – Weight of the factor, Fi – Value of the factor

A couple of questions arise. First, how do we determine the weight associated to each factor and second, how is the value of each factor calculated? Table I gives a brief idea for computation of values of each factor:

Table I. Ratios taken over a period T

Sno.	Factor Name	Computaion formula
1	Posting Time	(t' – t1)/(t2 – t1) t1 – last login t2 – current login t' – time of the post
2	Posting	(Post on the poster's timeline)/(total posts)
3	Likes	(total likes on poster's posts)/(total likes)
4	Comments	(total comments on poster's posts)/(total comments)
5	Shares	(total shares of poster's posts)/(total shares)
6	Tags	(total tags in poster's post)/(total tags)
7	Post popularity	(like on poster's post)/(likes on all posts)
8	Grouping	Assign the group value
		Group Value = ∑Fi(group)/n where n=7 i.e. total number of factors

Table II. User feedback form

Factor Name	Very important	Important	Should be considered	Not so important	Least important
Posting Time	X				
Posting	X				
Likes		X			
Comments			X		
Shares		X			
Post Popularity			X		
Grouping				X	
Tags		X			

Table III. User feedback form

Factor Name	User 1	User 2	User 3	User 4	Mean Value
Posting Time	5	4	3	5	4.25
Posting	4	4	3	3	3.5
Likes	3	4	3	4	3.5
Comments	4	4	4	3	3.75
Shares	3	3	3	3	3
Post popularity	2	3	2	3	2.5
Grouping	3	3	3	3	3
Tags	3	4	5	3	3.75

Now, to find the weights associated to each of the factor we use the technique of Linear Regression. We conduct a survey wherin we tell people to assign weights to each of the factors. For this, we use a four level rubric scale (Table II). After obtaining user feedback, we calculate mean value for each factor (Table III). The mean value of each factor gives the popular perception of their weights. Now for the first few posts the priority number is calculated by adding the mean values Mi for all the factors Fi which apply to that post. For example, suppose you liked the post but didn't share or comment then the value of F(p) would be just Mlike. However, if you like, comment and share the post F(p)=Mlike + Mcomment + Mshare. Also we obtain the values if Fi during this time. Once we get the minimum number of equations, we find the all the values of Wi in the following way:

$$F(p)_1 = w_1F_1 + w_2F_2 + w_3F_3 + w_4F_4 + w_5F_5 + w_6F_6 + w_7F_7 + w_8F8$$

$$F(p)_2 = w_1F_1 + w_2F_2 + w_3F_3 + w_4F_4 + w_5F_5 + w_6F_6 + w_7F_7 + w_8F8$$

$$F(p)_3 = w_1F_1 + w_2F_2 + w_3F_3 + w_4F_4 + w_5F_5 + w_6F_6 + w_7F_7 + w_8F8$$

…

$$F(p)_8 = w_1F_1 + w_2F_2 + w_3F_3 + w_4F_4 + w_5F_5 + w_6F_6 + w_7F_7 + w_8F8$$

Once we obtain the weights of all the factors, we calculate the value of F(p) by using $F(p)=\sum w_i F_i$. We no more calculate F(p) using the M_i since we now have the weight for each factor. Consequently, each post is associated with a priority number. The greater the value of the priority number the higher it's priority and will appear at the top in the news feed. Similar mechanism can be used to sort news articles and posts from pages.

3. EVALUATION

The article [3] throws the question "Are You on Social Overload?" at the internet users. A list of 16 frequently used social channels with distinct purposes and requiring distinct engaging rules are listed. The point made is that social media is becoming messy and requires strategic time management. Among the list, Facebook, launched in 2004, became the largest social networking site in the world in early 2009 [4]. Facebook was first brought to the public with its attempt to make the world more open and connected". [6] shows some interesting Facebook user statistics. According to the stat the average time spent per Facebook visit is 20 minutes. On average each time a login user is exposed to 1500 potential pieces of content. The mass of information on Facebook consists of 3 categories namely stories and photos of friends, news feed and posts of public pages. Users are overwhelmed partially because the posts from all categories are displayed together without any particular sequence. Our solution addresses this issue by displaying the posts under three different categories. In this way the user can focus on a specific category without being distracted.

Moreover, a study on college-aged adults carried by University of Michigan [5] finds that the more they used Facebook the worse they felt. The study suggests that in social interactions users get more aware of what their friends are doing even though that may have nothing to do with their own life. Meanwhile, a lot of Facebook users share the same experience as they login their account, scroll down the pages until they find something exciting at the bottom of the page. Useless information gets flooded onto the user before they get what they want. The overload could be eliminated by displaying posts in a different order so that the user interested ones are placed on the top. To achieve this goal the solution applies mathematical functions to measure the priority of each post according to user's preference shown in past behavior. The three metrics used for priority measurement are interest of the owner for each post, popularity of the post and posting time. Firstly we measure user's interest in each of his friends by collecting the number of comments that the user placed to his friend's posts as well as his posts on each friend's timeline. The second metric popularity of each post is indicated by the total number of likes. The last but not the least, we take update time into account to filter out the stale news since we hope to give higher priority to the stories which have been posted closer to the time of user login. The weights assigned to all factors are dynamically adjusted in the way that tends to give the most interesting post the highest priority.

Finally we compare the algorithm in our solution with the ranking algorithm used by Reddit, which is widely accepted as the closest to produce quality online discussions. The algorithm [1] assigns a score to each story based on the post time as well as the up votes and down votes. The latter two factors measure the popularity of a post which share similar idea with our solution. The comment ranking algorithm purely takes the rating of the comment into account while the update time has no effect on comment ranking. Similarly in our solution the measurement of user's interest in each of his friends can be regarded as a personal rating to all friends.

4. CONCLUSION AND FUTURE WORK

The consequence brought by social media overload is that people have to accelerate the rate at which they consume the news and in return losing time for meaningful communication and discussion. When we look into Facebook the slogan on sign up page has changed from "connect and share with the people in your life" to "stay connected", which loosened the connection with our real life. The mixed form of displaying all stories from friends, news, public pages and even ads just simply places everything in front of the user's face. Only two sorting strategies namely most recent and top stories are applied to determine the order of all posts regardless of the user's interest. The solution we proposed addresses both issues by categorizing different time of news and put the stories into the order regarding user's interest. The algorithm gives the news with the latter update time, higher popularity and posted by more interesting friends a higher priority on the list. With the carefully ordered post list the user will always read the news they are more likely to be interested in on the top of the page. As the page goes down the user can figure out the posts are not attracting or irrelevant to their life and then put them aside. In this way users will suffer less from the mass of meaningless news posted on their Facebook page. As our priority measurement algorithm is specially designed for sorting Facebook friends' news. It is not likely to be directly applied to other scenarios. The future work lies in adjusting the algorithm to the other two categories, news and public page posts. In addition as we take users' personal interest in their friends into consideration the variety of the top posts is no longer guaranteed. The resource of the posts showing on top of the pages will be narrowed down to a small portion of all the user's friends. In the future we need to find a balance between reducing the large amount of irrelevant and unattractive stories and enabling the variety of the stories.

REFERENCES

[1] How reddit ranking algorithms work. http://amix.dk/blog/post/19588, November 2010.

[2] Aaron Elliott. Information overload from social media. http://socialmediatoday.com/socialbarrel/1637856/information-overload-social-media-infographic, August 2013.

[3] Peter Himler. Are you on social overload? http://www.forbes.com/sites/peterhimler/2013/01/30/are-you-on-social-overload/, January 2013.

[4] Andy Kazeniac. Social networks: Facebook takes over top spot, twitter climbs. https://blog.compete.com/2009/02/09/facebook-myspace-twitter-social-network/, February 2009.

[5] Ethan Kross, Philippe Verduyn, Emre Demiralp, Jiyoung Park, David Seungjae Lee, Natalie Lin, Holly Shablack, John Jonides, and Oscar Ybarra. Facebook use predicts declines in subjective well-being in young adults. PloS one, 8(8):e69841, 2013.

[6] Craig Smith. By the numbers: 75 amazing facebook statistics. http://expandedramblings.com/index.php/by-the-numbers-17-amazing-facebook-stats/, January 2014