School of Computer Science and Engineering

Faculty of Engineering

The University of New South Wales

Managing Your Social Networking Profile

Enabling User-Tailored Views of Your Feed

by

Geoffrey Zhu and Richard Zhang

Thesis submitted as a requirement for the degree of

Bachelor of Engineering in Computer Engineering

Submitted:	26th May 2015
Supervisor:	Dr Hye-young Paik
Student ID:	3415440 & 3416477
Topic ID:	3031

Contents

Chapter 1

Introduction

Social Networking Services (SNS) are platforms where a diverse range of users are able share their interests, organise social activities and keep in touch with people. In almost all SNS, users are presented with a feed; this feed is a list of items generated via the user's connections that the user may be more or less interested in. The feed acts as a summary of activities that the user has subscribed to, a dashboard presented to them when they first log in. This feed contains a large amount of items that we would like to order, or rank in some way such that the items that the user finds more interesting have a higher precedence in the feed. As the use of SNS grows rapidly, so does the demand for such ranking algorithms.

While there exists many SNS currently being used, the scope of this thesis will be reduced to focus on only one of them; Facebook. This is mainly due to the time constraints involved, however the results and methodology used in both our research and implementation will be generalisable to all SNS. There are two main reasons why Facebook has been chosen as our SNS of focus. Firstly, Facebook is currently the most used SNS which allows us to more easily gather users for our purposes aswell as have more confidence that our results can be generalised to most SNS. Secondly, Facebook attracts many different types of users due to it's very flexible, generic nature (i.e. not a niche SNS). Each type of user will have different wants and needs and by having a

large set of user types, we are able to more easily identify them.

With all this considered, it is clearly not possible to have one single ranking algorithm to accomodate for all users, and yet as of now, Facebook only offers one ranking, or view of a user's feed (besides chronological order). Our aim can be summarised as follows: Firstly we will set out to identify these different user types and their needs, then we aim to create a number of different ranking algorithms based on the discovered user types. Thus offering a more personalised ranking of a user's feed. It is important to note that we do not aim to create a *better* ranking algorithm than Facebook as an enormous amount of research and time has already been put into creating said algorithm, instead we aim to offer different, more personalised rankings.

Chapter 2

Background

2.1 User Modelling

- Discuss user modelling here
- Refer to papers and quote ...

Why do people use Facebook? I would like to cite Bob [NH12] who has died for no reason.

Four approaches to user modellinga qualitative research interview study of HCI professionals' practice I would like to cite Bob [Cle04] who has died for no reason.

MySpace and Facebook: Identifying dimensions of uses and gratifications for friend networking sites I would like to cite Bob [BRR10] who has died for no reason.

Semantic modelling of user interests based on cross-folksonomy analysis I would like to cite Bob [SAC+08] who has died for no reason.

2.2 Ranking Algorithms

A ranking algorithm will give each item a score and order them with the item with the highest score at the top and the item with the lowest score at the bottom. The score of an item will depend on a set of criteria that the ranking algorithm uses. Social networking services will use these ranking algorithms in order to provide the user with items that will interest them. Since we are focusing on Facebook, the users will be provided with a feed that contains a lot of posts that they will receive. In Li [LTL+10] paper, they discovered that there are three major factors that could affect how interesting a user may find a particular item. They are:

- Topical Preference
- Topological Locality
- Social Influence

Topical reference is the idea that most users are interested in a limited range of topics. Topological locality refers to the fact that users are interested in the topics that their friends like and Social influence basically says that users are interested in famous people such as singers or actors. These factors do provide some insight on what a user likes but could be further generalised to topics that a user likes. The Topological locality does raise an interesting notion, that is, users are more likely to like a post that a friend likes. We can summarize these two into a more general categorization of Topic classification and connections. Topic classification will be classifying the topics that a user may like while connections will be a measure of how 'close' the user is with their friend based on their interactivity.

Topic classification is quite difficult we have to generalise a topic that they may like based on the posts that they receive. In a paper by Bur [BMAC13], they analysed twitter tweets and tried to generalise a topic based on the tweets each user received. They have analysed two types of methods. They are:

- BestOverlap
- UserInfoBigram

The BestOverlap method attempts to gather a huge amount of tweets and look at the common words in those tweets. A topic can be generated by the word is overlapped the most across all the tweets. In regards to our algorithm where we have to look at Facebook posts, the likelihood of word overlaps across a large amount of posts is quite low. This method would not be appropriate for our purposes.

The UserInfoBigram analyses the optional text that is provided in every tweet and generalises a topic from those words. In Facebook, almost no one uses the optional text so this method will also not work.

In order to do topic classificication we had to look to (insert your related work richard!! and explain)

Aga [ACG+14] discusses activity ranking for LinkedIn which is also a Social Networking Service. They discover two more factors that have a huge impact on whether the activity is deemed interesting or not. They used the measurement of CTR or click-through-rate which is the probability of a user clicking on the link to measure the appeal of an activity. An activity that was old had a low CTR compared to an activity that was new. It seems that the freshness of an activity or the time that the activity was made had to be taken into account in the ranking algorithm. Another factor was diversity. A huge drop in CTR was found when they gave users a repeated type of activity in their feed. We can surmise that freshness and diversity are key factors that must be considered in our ranking algorithm. The method that they have used to deal with these two issues involved re-ranking the feed with a decay factor to account for time and adding a negative score to activities of the same type. For our algorithm, we plan to utilise the same methods proposed as they have been successful.

Aga [ACG+14] also reinforces the idea that people are interested in what their friends like when they analysed the activities of co-workers and colleagues. Like Li [LTL+10],

they found that there was an increase in the click-through-rate of activities they were made by co-workers in the same organisation and colleagues. This emphasizes the importance of the factor of connections.

Chapter 3

Plan

3.1 Proposal

Our thesis aims to provide a more personlized view of Facebook's feed that is more adaptable to users. We do this through the introduction of user types in order to figure out what users actually want in their feed. We utilize the same tried and true algorithms used in ranking the feed but we incorporate user types and the weights that are produced from this type in order to make the feed more relevant to the user. This means that users will be provided with posts that they are more interested in at the top of their feed.

3.2 Our System

Figure 4.1 provides a general overview of our implementation plan. It is a block diagram of our system.

In our design, we will have a front end module which will be the website that is seen be the user. The website will have a logic screen for user authentication which will allow us to pull the data from their Facebook feed. This is the job of the puller module.

The user will be given a couple of options of which user type that they think they are. The associated weights from the user types will be brought into the ranking module which contains two algorithms, one for topic classification and one for connections. The puller module will pull the feed in data into this ranking module which will rank the feed using the total scores from the topic classification, connection module and the assigned weights of the user types that came from the website or front end. To do topic classification we will look at the user's posts and try to generalise the topic based on what they have written. A method to do this is (insert text here)

The connections module will simply look at how often the user has interacted with the person who posted that item and give a score based on that. After the scores have been assigned to each post, the feed will be ranked on the score and passed over to the freshness module. We will use Aga's [ACG+14] method and assign a decay factor to feeds that are not as recent. This feed will be reranked based on the new scores and passed over to the diversity module. In this module, we will rerank the feed and add a negative score to consecutive posts of the same type. Lastly, the newly ranked feed will be displayed in the frontend in front of the user.

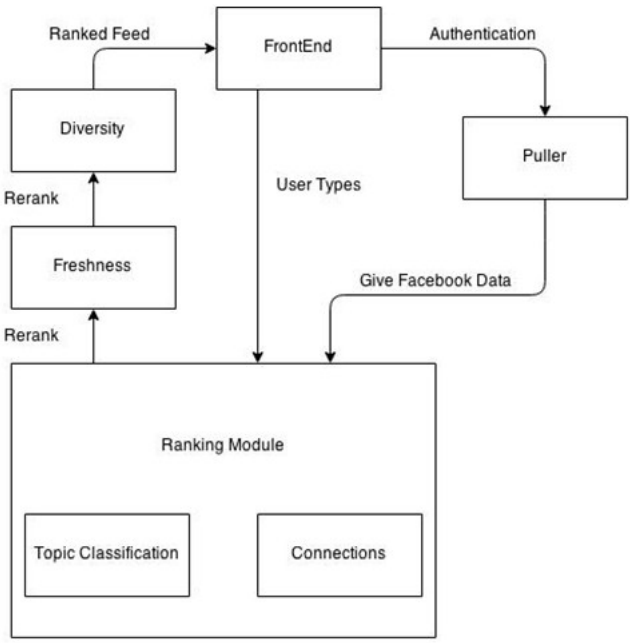

Figure 3.1: Block diagram

We plan to use nodejs to do the whole project due to its flexibility and adaptability. We will use some written nodejs API's in order to interact with the Facebook API. Our timeline will be is shown below.

3.3 Evaluation

We considered two different ways of evaluating our system.

The first method was to simulate real users by creating Facebook accounts and attempting to mimic behaviour of each user type. We would then create a ground truth regarding how that type of user would like their feed ranked and compare the output of our ranking algorithm to the ground truth. We found quite a few flaws in this method, the major one being how difficult it would be to simulate a real user. Creating social interactions and simulating connections between users would prove very difficult. On top of this, the ground truths that we would be creating could be affected by confirmation bias. This left us with a very questionable evaluation method, so we arrived at our second one.

The second, and chosen evaluation method takes the form of gathering real users and performing a form of usability test. In this test, we will ask participants to order their feeds how they would like it to be seen, this forms an unbiased ground truth. We then run our ranking algorithm on their feeds and compare the ground truth they gave us earlier to the output. In addition to this ground truth comparison, we will ask the user to compare our ranking algorithm with the one provided by Facebook, without telling them which is which. This will give us some subjective results as to whether our ranking algorithm has succeeded in personalising the user's feed.

Bibliography

[ACG+14] Deepak Agarwal, Bee-Chung Chen, Rupesh Gupta, Joshua Hartman, Qi He, Anand Iyer, Sumanth Kolar, Yiming Ma, Pannagadatta Shivaswamy, Ajit Singh, et al. Activity ranking in linkedin feed. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1603–1612. ACM, 2014.

[BMAC13] Matthew Burgess, Alessandra Mazzia, Eytan Adar, and Michael J Cafarella. Leveraging noisy lists for social feed ranking. In *ICWSM*, 2013.

[BRR10] Jennifer Bonds-Raacke and John Raacke. Myspace and facebook: Identifying dimensions of uses and gratifications for friend networking sites. *Individual Differences Research*, 8(1):27–33, 2010.

[Cle04] Torkil Clemmensen. Four approaches to user modelling-a qualitative research interview study of hci professionals' practice. *Interacting with computers*, 16(4):799–829, 2004.

[LTL+10] Huajing Li, Yuan Tian, Wang-Chien Lee, C Lee Giles, and Meng-Chang Chen. Personalized feed recommendation service for social networks. In *Social Computing (SocialCom), 2010 IEEE Second International Conference on*, pages 96–103. IEEE, 2010.

[NH12] Ashwini Nadkarni and Stefan G Hofmann. Why do people use facebook? *Personality and individual differences*, 52(3):243–249, 2012.

[SAC+08] Martin Szomszor, Harith Alani, Ivan Cantador, Kieron OHara, and Nigel Shadbolt. *Semantic modelling of user interests based on cross-folksonomy analysis*. Springer, 2008.