Predicting Engaging Content for Increasing Organic Reach on Facebook

Natthaphong Phuntusil and Yachai Limpiyakorn(✉)

Department of Computer Engineering, Chulalongkorn University, Bangkok 10330, Thailand
Natthaphong.Ph@student.chula.ac.th, Yachai.L@chula.ac.th

Abstract. Over the past few years, many people have been concerned about declines in organic reach for their Facebook Pages. This has been a pain for many businesses, especially those small businesses and startup. Organic reach refers to how many people you can reach for free on Facebook by posting to your page. The declined organic reach results from some key changes to improve how News Feed chooses content. News Feed is aimed at becoming more engaging, even as the amount of content being shared on Facebook continues to grow. This paper presents a technique to increase Facebook organic reach. The method investigates some promising factors to predict the engaging content posting on business Pages, so that the post would gain exposure in News Feed of the liking users on Facebook. The proposed approach provides the alternative for businesses to increase the organic reach without more expense on advertising posted on Facebook Pages.

Keywords: Social commerce · Decision support · Engaging content · Organic reach · Social network

1 Introduction

Facebook users have probably felt that more and more content is being created and shared on social media every day. Thanks to devices like smartphones, many people can share important moments and experiences, photos and videos, or articles with just a few swipes of the finger or taps on a button. In addition to the growth in content, people are also liking more Pages. As a result, competition in Facebook News Feed is increasing, resulting in harder for any story to gain exposure in News Feed.

News Feed is the place on Facebook where people view content from their family and friends, as well as businesses [1]. It is designed to show each person on Facebook the content that is most relevant to them. Some key changes to improve how News Feed chooses content have resulted in declined organic reach. Organic reach refers to how many people you can reach for free on Facebook by posting to your page [2]. Declines in Pages' organic reach have been a pain for social commerce. Many businesses, especially small businesses and startup, are concerned about the rising expense of advertising their business Pages. This research has thus investigated some factors potentially predict the engaging content posting on business Pages, so that the post would gain exposure in News Feed of the liking users on Facebook. Finally, this would result in the increase of organic reach on Facebook Pages, and less advertising cost for businesses.

K. Kim and N. Joukov (eds.), *Information Science and Applications 2017*,
Lecture Notes in Electrical Engineering 424, DOI 10.1007/978-981-10-4154-9_73

2 Facebook

2.1 Facebook Pages

Pages are for brands, businesses, organizations and public figures to create a presence on Facebook, whereas profiles represent individual people. Anyone with an account can create a Page or help manage one, if they have been given a role on the Page like admin or editor. People who like a Page and their friends can get updates in News. There are 6 primary categories [3–5] to choose from, including: Local Business or Place; Company, Organization or Institution; Brand or Product; Band or Public Figure; Entertainment; Cause or Community.

2.2 Facebook Graph API

The Graph API is the primary way to get data in and out of Facebook's platform. It is a low-level HTTP-based API that can be used to query data, post new stories, manage ads, upload photos and a variety of other tasks that an application might need to do [6].

3 Similarity Measurement

Term Frequency–Inverse Document Frequency, or TF-IDF, is a numerical statistic used as a weighting factor in information retrieval and text mining [7]. The intent is to reflect how important a word is to a document in a collection or corpus. Zhang and Pennacchiotti [8] studied the correlation between Facebook categories and eBay meta-categories. The result suggested that the set of Facebook categories may be predictive of purchase behaviors. The list of Facebook categories contains 214 features. For each user u and Facebook category f the feature value is computed using TF-IDF to reflect the user interest associated with a particular category.

4 Research Methodology

Figure 1 illustrates steps of the proposed method. In brevity, to predict the engaging content, the process starts with creating the selected business Page, then comparing the similarity of user interests between people found in the business Page and content page sources. Cosine similarity and Pearson correlation are applied for the similarity analysis in this work. It is observed that many null values exist in the dataset. However, it does not mean there is none of interests on null category. It might be pages not found in null category. Therefore, Pearson correlation is used to test the result by minus with the average for shifting the value to centre.

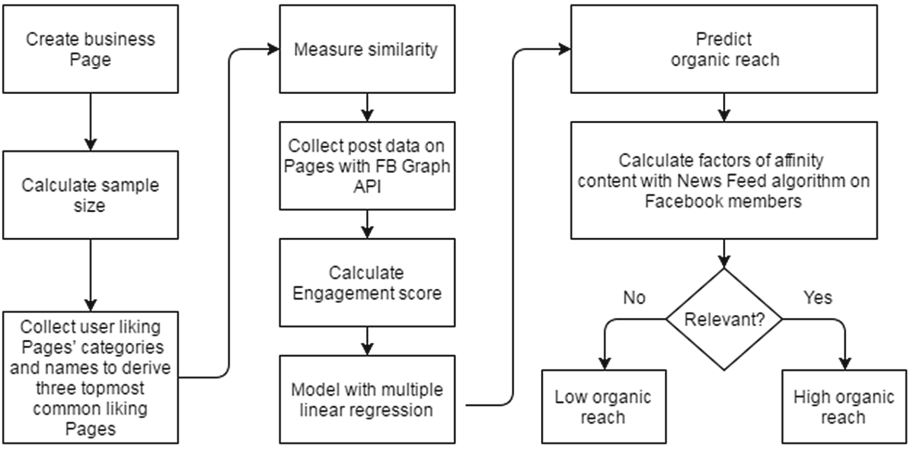


Fig. 1. Process of identifying engaging contents to increase Facebook organic reach

4.1 Create Business Page

For the experiment, a new business Page was created. The audience was built by promoting the Page with Facebook ads to the target goal of 1,000 fans (Page likers). Meanwhile, we had been posting to sell products on the business Page timeline. Figure 2 shows the findings reported by Facebook Insight. We found that 96% is Women and 4% is Men. The information suggested that the target group would be women. The post plan will then focus on the engaging content for women.

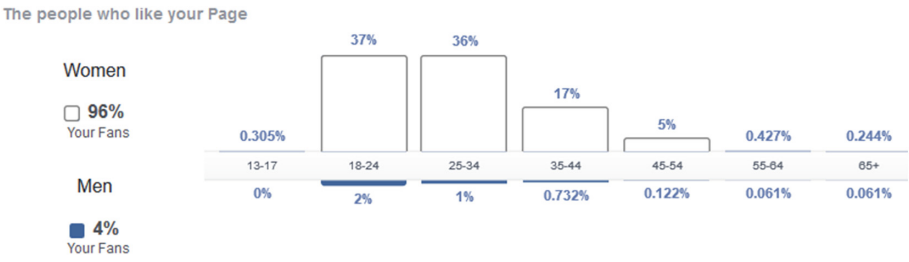


Fig. 2. Findings of created business page reported by Facebook insight

4.2 Calculate Sample Size

Based on Yamane [9] formula, the size of dataset containing 400 users is used as the sample size n in this work, while the population size $N = 1.18$ Billion daily active users. The result provides 95% confidence level, and level of precision $e = 0.05$.

4.3 Collect User Liking Pages' Categories and Names

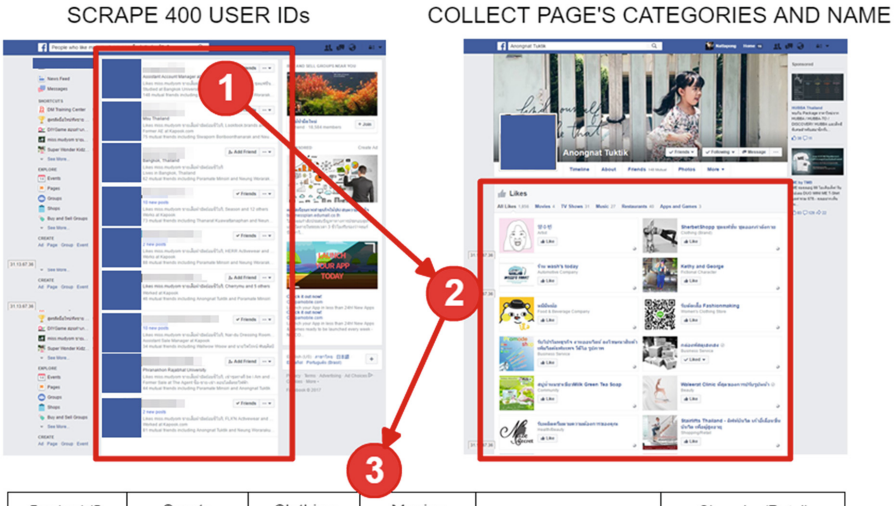
A set of 400 users liking the created business Page was randomly selected. The library of Python 3, Selenium Webdriver [10], is used to control Firefox browser to scrape the information of all 400 users liking FB pages' categories and names. This is the challenging step to collect the important data of the counts and categories of other FB liking Pages from the 400 users who likes the created business Page. Table 1 summarized the details of the top three liking FB Pages of the selected 400 users.

Table 1. Top three of liking FB pages collected from 400 users liking created business Page

FB page	Content type	Fans' likes	*Like(u, f)*[a]
1. Spice	Beauty, Healthy	686,563	197
2. Vonvon.me	Entertainment, Game	28,922,326	149
3. Jatiewpainai	Travel	788,509	145

[a]*Like(u, f)* is the number of page likes by user *u* in page *f*

Figure 3 shows the Facebook users' interests based on Pages they like. Although there are many reasons for users to like their Pages, we assume that people want to get more useful information from Facebook Pages. In this work, the amount of liking page



Facebook ID	Sports	Clothing	Movies	.....	Shopping/Retail
755805214	2	10	1	.....	45
1084190380	42	47	2	.....	33
.....	.....	.....	.....	.....	.....
617214819	12	457	78	.....	412

Fig. 3. Collected data of liking user Pages' categories

categories is used to measure similarity of user interests between the business Page and content page sources.

4.4 Measure Similarity

Currently, the list of Facebook categories contains 156 features. User page's categories in terms of frequency is used to calculate the similarity of people [11]. The sampling data of 400 users are used for the computation of cosine similarity between the business Page and sources of content page. If pairwise vectors are same orientation, then cosine similarity equals 1, whereas 0 denotes both vectors are not similar. The similarity threshold is set to 0.5 or 50% in this work. Example similarity results are:

Spice (41%), Jatiewpainai (32%), Starvingtime (43%), WomenBeautyCommunity (40%), SistaCafePage (38%), Jeban(30.5%).

4.5 Collect Post Data on Pages with Facebook Graph API

The post data on Page were collected from Facebook timeline to calculate engagement scores and post scores. The collected fields consist of: *"status_id", "status_message", "link_name", "status_type", "status_link", "status_published", "num_reaction" are summation of all user reactions such as "num_comments", "num_loves", "num_wows", "num_hahas", "num_sads", "num_angrys", "num_shares", "num_likes".*

4.6 Calculate Engagement Score

There are many Facebook engagement score formulas. Equation 1 shows the engagement score formula from the website unmetric.com. The formula is derived from user research and observations on features and functionalities of different social media platform [12, 13]. Table 2 describes the variables used in Eq. 1. The default unmetric values of α, β are used in the formula.

$$E_{score} = \frac{(N_{like} + \alpha N_{comment} + \beta N_{share})}{N_{audience}} \times 10^4 \tag{1}$$

Table 2. Variable description.

Symbols	Meaning
E_{score}	Engagement rate
N_{like}	Number of likes on for each posts
$N_{comment}$	Number of comments for each posts
N_{share}	Number of shares for each posts
$N_{audience}$	Audience reception rate = $(Number\ of\ brand\ fans)^{0.8}$
α	5
β	10

In this work, Share is considered as the factor with more influence on the audience reach, compared to Like and Comment. We focus on the highest shares of content page post and then calculate the engagement score for prediction. The result of extrapolation is not sensible scoring because the similarity of each page does not equal. The Similarity score is thus applied as the weight to adjust the engagement score in order to obtain the Post score as shown in Eq. 2.

$$Post_{Score} = E_{score} \times Sim_{score} \tag{2}$$

The Post score will then be used to pick the content for posting on the business Page. Next, the evaluation of the effect results of engagement scores between two pages is carried out. The content with the highest score from Facebook source pages was posted as well as a variety of contents posted in random time within the same day.

Figure 4 illustrates how to derive the values of $Sim_{score}, E_{score}, Post_{Score}$ in sequence.

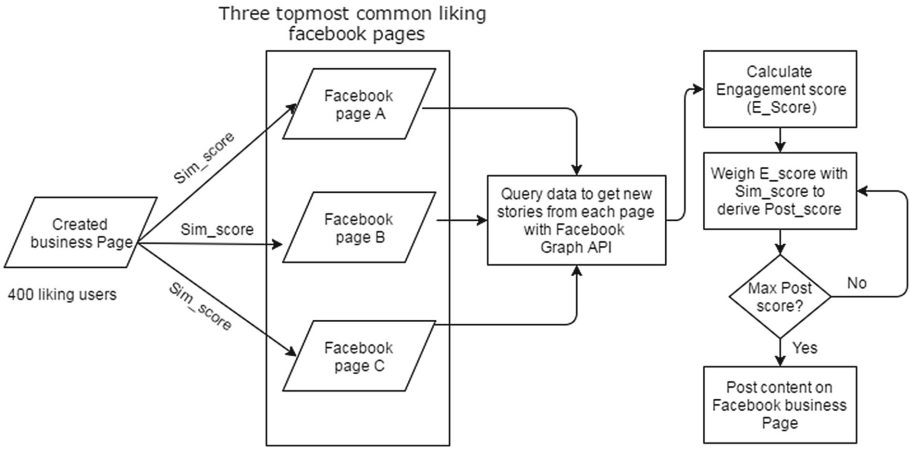


Fig. 4. Steps to derive values of similarity score, engagement score, and post score

4.7 Multiple Linear Regression

Multiple Linear Regression as Eq. 3 is used to predict $\hat{Y}$ on the basis of p predictors $(X_1, X_1, \ldots, X_p)$.

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + \ldots + b_pX_p \tag{3}$$

The initial set of 12 predictors consisting of:

$E_{score}, Sim_{score}, Post_{Score}, Reactions, Comments, Shares, Likes, Love, Wows, Hahas, Sads, Angrys$

was used as independent variables to predict organic reach using multiple linear regression. Throughout several iterations of adjustment, we obtain the best regression model with three independent variables: (1) Similarity scores, (2) Shares, and (3) Wows, used as the factors to predict the value of dependent variable, organic reach, in this work.

5 Results

The highest twenty E-score engaging contents were posted on the created business Page in each period of content page sources. It is observed that some contents can penetrate the reach baseline. This may assume that the posts could be considered as affinity contents. The multiple correlation between reach and the three predictors = 0.9246 and $R^2 = 0.8549$ or 85.49%. However, we are more interested in Adjusted $R^2 = 0.779$, that is 77. 9% of the variation in organic reach can be predicted on the basis of the three predictors. The value of standard error is 130.874.

The hypothesis was established as below:

H_0: If there is no relationship between organic reach and selected factors.

H_1: If there is relationship between organic reach and selected factors.

According to Table 3, where $\alpha = 0.05$; then $p - value = 4.05\text{E} - 07 < 0.05$,

Table 3. ANOVA Summary

Source	*df*	*SS*	*MS*	*F*	Significance F
Regression	3	1715860.74	571953.58	33.392	4.05E-07
Residual	17	291176.25	17128.01		
Total	20	2007037			

H_0 is rejected because *significance F* is less than the predetermined value 0.05.

We found that all the p-values of each independent variable were significant: SimScore = 0.000941, Shares = 0.00001713, Wows = 0.0001152. The regression coefficients can be used for weighing three predictors, that is, let b_0 represent the intercept = 0, $b_1 = 387.771, b_2 = 0.3759, b_3 = -7.818$.

Equation 4 shows the model derived for supporting the prediction of organic reach on Facebook with the use of three factors: Similarity score, Shares, and Wows.

$$\hat{Y} = 387.771\, SimScore + 0.3759\, Shares - 7.818 Wows \tag{4}$$

6 Conclusion and Future Work

The News Feed ranking system offers people a better, more engaging experience on Facebook. However, the policy has affected social commerce since it causes Pages' organic reach to decrease. This paper presents an approach to supporting the decision for post plan. Rather than spending on more advertising, small businesses and startup could selectively post contents on their business Pages to increase Facebook organic reach, so that businesses can succeed on Facebook. We present a method to predict the engaging content for increasing organic reach on Facebook. A set of 12 factors was investigated whether they are potential predictors for engaging content. With multiple linear regression, the preliminary results on the clothing category showed that there is relationship between Facebook organic reach and the selected factors, which are Similarity score, Share, and Wow. The findings support the assumption that posting the viral affinity content could raise organic reach on Facebook. This would alleviate pain on

social commerce due to the policy of Facebook News Feed. However, further investigation of the relationship type, the potential predictors associated with individual coefficient need be carried out so that the output model could be extrapolated to other similar domain business Pages.

References

1. Facebook for Business. https://www.facebook.com/business/news/Organic-Reach-on-Facebook. Accessed 11 Sep 2016
2. An Update to Facebook News Feed: What it Means for Business Pages Facebook for Business. https://www.facebook.com/business/news/update-to-facebook-news-feed. Accessed 01 Dec 2016
3. Facebook Help Center. https://www.facebook.com/help/282489752085908. Accessed 11 Sep 2016
4. List of Facebook Fan Page Categories. http://www.birdsonganalytics.com/list-of-facebook-fan-page-categories/. Accessed 26 Nov 2016
5. Create a Page. https://www.facebook.com/pages/create/. Accessed 27 Nov 2016
6. Overview - Graph API - Documentation - Facebook for Developers. Facebook Developers. https://developers.facebook.com/docs/graph-api/overview. Accessed 12 Sep 2016
7. Karen Spärck Jones – Wikipedia. https://en.wikipedia.org/wiki/Karen_Sp%C3%A4rck_Jones. Accessed 20 Nov 2016
8. Zhang, Y., Pennacchiotti, M.: Predicting purchase behaviors from social media, pp. 1521–1532 (2013)
9. Yamane, T.: Statistics: An Introductory Analysis. Harper and Row, New York (1967)
10. Introduction — Selenium Documentation. http://www.seleniumhq.org/docs/01_introducing_selenium.jsp#introducing-selenium. Accessed 24 Oct 2016
11. Beginners Guide to Learn About Content Based Recommender Engines, Analytics Vidhya. 11 Aug 2015
12. Engagement - Unmetric. https://unmetric.com/engagement/. Accessed 07 Nov 2016
13. Bernstein, M.S., Bakshy, E., Burke, M., Karrer, B.: Quantifying the invisible audience in social networks. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 21–30 (2013)