Political attitude estimation through Facebook like: a South Korean case study

Muhammad Eka Wijaya, Meer Sadeq Billah & Heejune Ahn

To cite this article: Muhammad Eka Wijaya, Meer Sadeq Billah & Heejune Ahn (2017): Political attitude estimation through Facebook like: a South Korean case study, Asian Journal of Political Science, DOI: 10.1080/02185377.2017.1402357

To link to this article: http://dx.doi.org/10.1080/02185377.2017.1402357

Published online: 14 Nov 2017.
Political attitude estimation through Facebook like: a South Korean case study

Muhammad Eka Wijaya, Meer Sadeq Billah and Heejune Ahn

Department Electrical Information Engineering, Seoul National University of Science and Technology, Seoul, Republic of Korea

ABSTRACT
Recently numerous studies are conducted to estimate the human personality from the online social activities. This paper develops a comprehensive model for political attitude estimation leveraging the Facebook Like information of the users. We designed a Facebook Crawler that efficiently collects data overcoming the difficulties in crawling Ajax enabled Facebook pages. We show that the category level selection can reduce the data analysis complexity utilizing the sparsity of the huge like-attitude matrix. In the Korean Facebook users’ context, only 28 criteria (3% of the total) can estimate the political polarity of the user with high accuracy (AUC of 0.82).

KEYWORDS
Social Network Data Mining; Political Attitude Estimation; Facebook Likes; SNS; Asian Politics; Matrix Reduction

1. Introduction
Estimating political attitude of a person is a crucial factor in building human relation and political campaign. Personality traits, personal values, biography, background contextual factors have been studied and considered as means of estimation measures (Caprara, Schwartz, Capanpa, Vecchione, & Barbanelli, 2006; Gerber, Huber, Doherty, Dowling, & Ha, 2010). Caprara et al. (2006) reported one can classify voting result in Italian elections with 67.2% accuracy combining 10 personal traits and 10 values. Gerber et al. (2010) also obtained similar results but the effects on political decision of personality can depend on contextual factors such as races and education levels. However, almost all of the previous studies use self-reports, which has many limitations including but not limited to higher cost of large scale surveys, bias in survey participants, and delay in data collection. Before the introduction of ICT technologies, for example, Internet and mobile phones, the only means to obtain the personal behaviours and attributes was the paper based surveys. The paper based self-report approach demands considerably huge resource overhead in terms of time and expense and is often limited to the range of accessible people.

Social Networking Sites (SNS), introduced in the early 2000s, have been growing rapidly in diversity of services and popularity among people. ‘Online social networking sites’ is a web service that allows users to construct public or semi-public profile within a bounded system, articulate a list of other users with whom they share connection, and manipulate their list of connections and those made by others within the system.
(Boyd & Ellison, 2007). Social networking sites are popular among people for the online activity features they provide and 59% of the Internet users are reported to use at least one SNS (Hampton, 2015). A considerable number of people of all age groups are found to be engaged in political activities on the social networks (Conroy, Feezell, & Guerrero, 2012; Gulati & Williams, 2013; Quintelier & Theocharis, 2013).

Among many SNS namely Myspace, Twitter, LinkedIn, and Facebook, Facebook has been drawing most intensive interest in social science research ranging from law, economics, sociology, and psychology to management, marketing and politics. The vantage of Facebook as a data source in research results from two factors. Firstly, Facebook is the most popular social networking site. Of all social networking site users, 92% use Facebook (Hampton, 2015). As reported on Facebook company statistics (Facebook Inc., 2015), there were 936 million daily active users in total of whom 798 million were mobile users on average on March 2015. Specifically in South Korea, users who access Facebook at least once in a month are 11 million in number on an average since June 2013, which covers almost 25% of the Korean population. In contrast, the next most popular SNS, Myspace is only used by 29% of all SNS users (Hampton, 2015). Moreover, Facebook provides much diverse opportunities to access human behaviours such as connection to others, preference expression, and personal status information. This is another role playing factor in choosing Facebook as the basis of research in diverse domains.

There are reasons why we need to improve our skills of utilizing SNS for the analysis of public attitudes. Social scientists are sometimes accused of failing to examine actual behaviour, relying instead on hypothetical or retrospective self-reports of behaviour (Baumeister, Vohs, & Funder, 2007); behavioural residue left on Facebook provides a compelling source of measurable behaviour traces. It is useful to think of Facebook as an ongoing database of social activity with information being added in real time (Wilson, Gosling, & Graham, 2012).

Facebook provides many features to its users for active participation. This paper, focusing on Facebook Group and Likes, studies the political attitude estimation of the users through Likes. The contribution of the paper is two-folds, in engineering and social science. In engineering, we developed an integrated crawler procedure for Facebook political data collection, without any dependency to a Facebook application recruiting participants. Also, we developed a category-based Facebook Like selection method to reduce the computational load in massive large user-Like matrix. The full analysis, flowing from the category selection method and SVD (single value decomposition) to logistic regression, can be done in a few 10 minutes by a non-distributed big data processing software package like R on the general PC environment. In social science domain, only a few studies have been reported on Asian countries (Hsu & Park, 2012) whereas an increasing number of social studies on Facebook have been conducted recently in American and European countries (Halpern & Gibbs, 2013; Painter, 2015; Vesnic-Alujevic, 2012; Vitak et al., 2011). The reason behind this unbalanced study could be caused by Mainland China blocking Facebook (excluding Hong Kong and Macau) usage under its policy of Internet censorship (Websites blocked in mainland China, 2015). To our knowledge, this is the first study for applying Facebook information for estimating Asian countries, and the results show that Facebook based social study can also be as effective for this region as European and American cases.
The remainder of this paper is organized in different sections as follows. Section 2 covers the introduction of the related works on social science research using Facebook. The motivation of the paper and contribution are also discussed there. In Section 3, we described our data collection method, that is, Facebook crawling. The efficient complexity and dimension reduction method for the large collected data of Facebook Likes and political attitude estimation algorithm are described and demonstrated in Section 4. Finally in Section 5, we concluded the paper by discussing the implications and limitations of our approach and findings.

2. Related work

Facebook, incepted on 2004, is one of the most popular SNS right now. It focuses on connecting with friends, family, and acquaintances through sharing, chatting, and other features for interaction. The users are also able to express their personal preferences such as favourite movie, music, product, activities, and organization. Facebook’s features include (1) News Feed, (2) Friends, (3) Wall, (4) Timeline, (5) Likes, (6) Message, (7) Notifications, (8) Networks and (9) Groups. Among these, the study uses Facebook Likes feature because they provide items for almost all the categories and the data is simple binary type.

Facebook Likes is a mechanism used by Facebook users to express their positive association with (or ‘Like’) online content such as photos, status updates of friends, Facebook pages of products, sports, musicians, books, restaurants, or popular Web sites. Likes represent a very generic class of digital records, similar to Web search queries, Web browsing histories, and credit card purchases. Facebook Likes has nice properties for the user’s preference; it covers unlimitedly diverse domains yet being simple binary information that are easy for processing. To access (read and write) these information, one should use Facebook API through an application. The developer is required to register the application in Facebook developer site and get application id and secret key for releasing publicly.

In their intensive literature review on Facebook social study, Wilson et al. (2012) grouped the study topics into five categories: (1) ‘descriptive analysis of users’ (e.g. who is using Facebook and what are they doing while on Facebook?) 24%, (2) ‘motivation for using Facebook’ (e.g. why do people use Facebook?) 19%, (3) ‘identity presentation’ (e.g. how are people presenting themselves on Facebook?) 12%, (4) ‘role of Facebook in social interactions’ (e.g. how is Facebook affecting relationships among groups and individuals?) 27%, and (5) privacy and information disclosure (why are people disclosing personal information on Facebook despite potential risks?) 18%. The study topics are diverse and inexhaustible.

In political study, probably the most intensive topic on Facebook is its political impact (Kim, 2011; Kim & Johnson, 2006; Macafee, 2013; Pennington, Winfrey, Warner, & Kearney, 2015; Williams & Gulati, 2008). However, the opposite direction, the impact of political attitude on Facebook activities are less studied. Recently, Kosinski, Stillwell, and Graepel (2013) demonstrates that personal attributes from their ages, gender, political and religious view to their personality (Big5) can be predicted though Facebook Likes. They collected the data on US region through a Facebook survey app, called MyPersonality. Our study is motivated by their work. However, their approach has several limitations that discussed in later sections. Their data collection is done through a Facebook
application, which usually takes longer time than crawling to collect data. Also the collected data can be biased in sample statistics because the participation to the surveys is voluntary. Also in data analysis step, the size of Facebook Like pages for the Facebook users in collected data easily exceeds several hundred thousand. When data analysis algorithms such as regression model, PCA, and SVD are applied to the user-like matrix, the computational complexity can be too high for a conventional computing environment.

3. Data collection method

Before describing the data collection method, we would like to briefly introduce the Korean politics and political parties. Korean national assembly has a multi-party system. However, the ruling and major opposition parties of South Korea hold most seats (290 out of 298) in the 19th Assembly, lasting from 2012 to 2016. The Saenuri party holds a majority of 160 seats, followed by the Saegungchi Party (more formally New Politics Alliance for Democracy) securing 130 seats. The Saenuri Party (the ruling party) practices center-right and conservative political views while the Saegungchi Party is a social-liberal political party (Manyin, 2003; Politics of South Korea, 2015).

3.1. Political attitude sources in Facebook

The political attitude data of Facebook users can be found in four major locations in Facebook: ‘political views’ in profile pages, being a friend of or following a politician, liking a (political) Facebook page, and joining (political) Facebook group. Figure 1 shows the

![Figure 1](image-url)
corresponding components in Facebook web pages. Table 1 shows the roles/definition and the trade-off in Facebook API support, range of information accessibility, information sources of the Facebook political sources.

Firstly, users can (optional features) declare their ‘political views’ in their profile sections. However, quite a small number of users shared their political views in the about section of Facebook. We observed less than 8% of US Facebook users from MyPersonality data and less than 3% Korean Facebook users from our collected data fill and share their political views. This low ratio is considered because political views are regarded as sensitive information open to public. The second source of getting political attitude is the list of followers and friends of a Politician. Facebook provides API for a friends and followers list of a Facebook user, but the crawler need an extended permission from the Facebook user. Also being a friend or follower to a specific politician cannot always guarantee the political support of the user to the corresponding party. The third source is political Facebook pages such as politicians’ Facebook pages. Facebook users express their political attitude by liking the politics related page and getting any updates from that page. Unfortunately, Facebook does not provide API to get the likers list of a Facebook page. It would be possible, but still difficult, if we randomly choose users and check the pages that they like. The last source is political Facebook groups that we used as the source for our data analysis. Facebook users join a community group to discuss and share their opinions. Becoming a member of a politics related group can be inferred as political tendency of a user related to that group. Facebook also provides publicly accessible API to get group members. The paper uses this Facebook group as political attributes sources. Joining a group can encourage political participation similar to offline group and it also produces same effects similar to offline group membership (Conroy et al., 2012).

### 3.2. Facebook crawler

Data in Facebook can be collected online in three major ways: recruitment of participants via Facebook applications (Kosinski et al., 2013), data crawling (Catanese, De Meo, Ferrara, Fiumara, & Provetti, 2011; Gjoka, Kurant, Butts, & Markopoulou, 2011), and combining Facebook applications and data crawling (van Dam & van de Vekdeeb, 2015). As the app-based recruiting method takes longer time for recruiting participants and the recruited people can be biased in statistics, we chose crawling method to collect political and Facebook Like data. The Facebook crawler used in the study is an integration of Scrapy framework (Scrapy Architecture, 2015) with Selenium (Selenium Introduction, 2015).

<table>
<thead>
<tr>
<th>Sources</th>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profile section</td>
<td>Most explicit.</td>
<td>Small number of users.</td>
</tr>
<tr>
<td>Facebook group</td>
<td>Most of the group info is publicly shared.</td>
<td>The number of users who joined in a group is smaller but greater than those who explicitly declare attitude.</td>
</tr>
<tr>
<td>Liking politician’s pages</td>
<td>Number of likers is big.</td>
<td>No direct API to get the likers.</td>
</tr>
<tr>
<td>Being a friend of follower</td>
<td>Can access the members via API</td>
<td>Available from like section of each user.</td>
</tr>
<tr>
<td></td>
<td>Some politicians publicly shared follower list with large number of followers.</td>
<td>Need extended permission to the users.</td>
</tr>
</tbody>
</table>

Table 1. Comparison of political attitude sources in Facebook.
Scrappy framework is an extensible open source web-crawler software, and provides concurrent HTTP requests and extension API for user-defined processing modules, called Spider. Selenium is an extensible web-browser. Selenium is combined with the Scrappy-based crawler to execute JavaScript and download HTML text, because Facebook page consists of multiple Ajax based java script components for each sub function.

Figure 2 illustrates our three-tier architecture Facebook crawler. The backend tier is data storage for storing crawled and extracted information. The middle tier is the Scrappy framework and Spiders that handle crawling procedure and the front tier is the selenium browser that is used for simulating user actions on Facebook. Facebook access token should be used to use the Facebook API. Basically, a user can get a token by generating it on Facebook Graph tools explorer. But the lifespan of a token is short (several hours) for our crawling period. We registered a Facebook application, combined it with short term token from Facebook Graph tools explorer, and then used the generated token to authenticate a crawler. Facebook checks received request patterns from a web browser or a Facebook application in order to prevent disruption in its service due to possible attacks like the request flooding. Facebook also checks access behaviour pattern. For example, accessing profile sections of other users randomly will be considered as a suspicious action. Detection of any suspicious request results in an access suspension for a specific time span according to Facebook terms of service. To avoid access suspension, the crawler uses 5-second interval between successive requests, and switches its browser name randomly in its HTML request. The detailed procedure for crawling is performed in the following order:

(a) Political Facebook Group List: The Korean political Facebook group is obtained from query using two Korean major political party’s names (Saenuri and Sajungchi) in Table 2.
(b) Group member list: Facebook crawling procedure starts from crawling group members using Facebook API. Facebook API provides public access to get Facebook group information.
(c) User’s Profile Page: In order to get users’ profile information, author must be logged in to simulate user action on Facebook pages. Selenium emulates user actions. First
information that is retrieved from this crawler is username or user id. Username or user id can be obtained by checking the redirected URL after viewing the user's home page. Other information such as education, gender, and location can be obtained from Facebook About section.

(d) User’s Like Page Data: Even though retrieving the pages information of a user that she liked seems to be possible to run simultaneously while crawling her profile, this step requires to be separated into different crawling process. Otherwise it is more likely that Facebook will detect the crawler as a robot for abnormal activity, the Facebook account used for crawling will be banned for several days as well. User’s Like information can be obtained without logging in Facebook. So, all Liked pages can be obtained by accessing directly to Facebook user’s pages without logging-in. After Facebook Like pages are retrieved, the crawler also gets page information such as Facebook page category, the number of Likes, and so on.

### 3.3. Collected data statistics

Facebook groups are of two types, either publicly or privately accessible. The authors collected data of group members from publicly accessible groups only as shown in Table 2. The users with less than five Likes are not included for data analysis. The collected dataset is from 1332 Facebook users in total of which 651 and 681 are supporters of Saenuri and Saenuri respectively; 725 users are male and 146 are female, and 105 users did not disclose their gender information. The gender imbalance would be due to the imbalance of political interest between genders of Korean population. From collected data, average Facebook page Likes per user (shown in Figure 3) is 262 with standard deviation value of 644 implying a large deviation in Facebook Like activity.

### 4. Political attitude estimation and experimental result

The authors developed a customized classification algorithm based on Logistic regression for estimating Facebook users’ political attitude from their Facebook Likes.

$$a_u \sim l_{u,1} + l_{u,2} + \cdots + l_{u,M},$$  \hspace{1cm} (1)
where $a_u$ is political attitude for a user $u$ (1, 2, …, N)

$$a_u = \begin{cases} 1 & \text{for a ruling party supporter} \\ 0 & \text{for an opposition party supporter} \end{cases}$$ (2)

and for page index $p = 1, 2, …, M$,

$$l_{u,p} = \begin{cases} 1 & \text{user u likes page p} \\ 0 & \text{otherwise} \end{cases}$$ (3)

The collected dataset can be represented in a ‘user-like’ matrix (Table 3):

$$L = \begin{bmatrix} l_{1,1} & \cdots & l_{1,M} \\ \vdots & \ddots & \vdots \\ l_{N,1} & \cdots & l_{N,M} \end{bmatrix}$$ (4)

Due to the huge number of users and Like pages (in this study, $1332 \times 95,675$), matrix dimension reduction techniques were essential for data analysis computation. Kosinski et al. (2013) used SVD algorithm for dimension reduction, but they apply the algorithm directly on the user-like matrix. The SVD algorithm provides two benefits in big data regression modelling, matrix dimension reduction and multi-collinearity remedy. However, the authors found directly applying SVD algorithm on the user-Like matrix takes too long time if not impossible in conventional statistical analysis tools such as R and MATLAB. Therefore in this paper we used two step approaches of Categories and Likes.

Table 3. User-like matrix example.

<table>
<thead>
<tr>
<th>Facebook ID</th>
<th>Chu S.\textsuperscript{a}</th>
<th>Kim Y.\textsuperscript{b}</th>
<th>Lee J.\textsuperscript{c}</th>
<th>For PGd\textsuperscript{d}</th>
<th>Supporting party</th>
</tr>
</thead>
<tbody>
<tr>
<td>98501288485XXX</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Saenuri</td>
</tr>
<tr>
<td>163998326955XXX</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Saenuri</td>
</tr>
<tr>
<td>82561944419XXX</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Saenuri</td>
</tr>
<tr>
<td>77116832633XXX</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Saenuri</td>
</tr>
<tr>
<td>64666071213XXX</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Saejungchi</td>
</tr>
<tr>
<td>77988187211XXX</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Saejungchi</td>
</tr>
<tr>
<td>95773477760XXX</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Saejungchi</td>
</tr>
<tr>
<td>10000132157XXX</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Saejungchi</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Chu Shinsu: Korean US major league baseball [MLB] player.

\textsuperscript{b}Kim Yuna: Famous Olympic gold medalist.

\textsuperscript{c}Lee Jaesu: Citizen is answer.

\textsuperscript{d}For President Geunhae Park.

Figure 3. Histogram of the numbers of Facebook likes of crawled users.
Figures 4 and 5 illustrate the proposed two step algorithm of category selections and Like-political attitude estimation, respectively.

4.1. User-category matrix

The authors noted that a Facebook page has its category depending on its subject. For instance, the current US president Barack Obama’s and Korean president Park Geun-Hye’s pages are in politician category and pages of Kim Hee-sun and Emily Watson are in actor/director category. Noticing that some categories have strong contribution to political attitude estimation but others do not, we can suppress the less correlated categories to reduce the matrix dimension dramatically. Denoting $c = 1, 2, \ldots, K$ for category index and the association of pages to a category $C_k = \{p(c)\}$, the user-category matrix then can be expressed in

$$L^C = \begin{bmatrix}
l_{c,1} & \cdots & l_{c,K} \\
l_{c,1} & \cdots & l_{c,K} \\
\vdots & \ddots & \vdots \\
l_{c,1} & \cdots & l_{c,K}
\end{bmatrix}
$$

An example user-category matrix representation (part of data set) is presented in Table 4. A user with id 98501288485XXX likes two Facebook pages in actor/director category, one page in athlete category, six pages in automobiles category, and nine pages in community category.

4.2. Multi-collinearity remedy

Multicollinearity often happens when the model has too many input variables, because the change that highly correlated variables included in the model increases with the increasing number of input variables. This case happens in the analysis of our study, because we cannot choose the input variables, which is determined from the Facebook pages.

Figure 4. Like page category selection analysis flow.
Therefore the next step is to check cross-correlations among categories. This step should be observed to avoid multi-collinearity problem before generating logistic regression model. In order to avoid multi-collinearity, first we checked the highly correlated categories. Checking the correlation coefficient among categories can be done as follows. First, find covariance between two categories and then standard deviation from two categories to solve the equation.

\[
    r_{X_1, X_2} = \frac{\text{Cov}(X_1, X_2)}{\text{STD}(X_1) \ast \text{STD}(X_2)}. \tag{6}
\]

When there is high correlation between two categories (the used threshold = 0.5), only one of them is selected for further analysis. In the collected dataset, the highly correlated

<table>
<thead>
<tr>
<th>Facebook ID</th>
<th>AC</th>
<th>AT</th>
<th>AU</th>
<th>CO</th>
<th>Related party</th>
</tr>
</thead>
<tbody>
<tr>
<td>98501288485XXX</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>9</td>
<td>Saenuri</td>
</tr>
<tr>
<td>163998326955XXX</td>
<td>12</td>
<td>4</td>
<td>18</td>
<td>42</td>
<td>Saenuri</td>
</tr>
<tr>
<td>82561944419XXXX</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>88</td>
<td>Saenuri</td>
</tr>
<tr>
<td>77116832633XXXX</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>Saenuri</td>
</tr>
<tr>
<td>64666071213XXXX</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Saejungchi</td>
</tr>
<tr>
<td>77988187211XXXX</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>18</td>
<td>Saejungchi</td>
</tr>
<tr>
<td>95773477760XXXX</td>
<td>0</td>
<td>2</td>
<td>10</td>
<td>35</td>
<td>Saejungchi</td>
</tr>
<tr>
<td>10000132157XXXX</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Saejungchi</td>
</tr>
</tbody>
</table>

Note: AC: actor; AT: athlete; AU: automobile; CO: community.
category examples are community/government category and cause category. Cause category is that relates with human life actions, such as ‘save the earth’ and ‘save the children’. Because both the community and the cause are similar, it is understandable that both of the categories are highly correlated. After multi-collinearity rejection, among 226 Facebook original categories are reduced into 109 categories.

4.3. Dimension reduction in user-category matrix

The main purpose of this step is to filter out some categories that have small correlation with political preference. Analysis method that will be used in this step is learning vector quantization (Rajaraman & Ullman, Mining of Massive Datasets, 2002). SVD cannot be applied in category because we want to keep the category without transforming its shape. Learning vector quantization, a branch of artificial neural network using a supervised method was used to improve regression decision regions.

By learning vector quantization, only 25 categories highly correlated with political preference remain. These 25 categories will be used in Facebook Page analysis. Top five correlated categories shown in Figure 6 are travel leisure, TV network, sport league, games toys, and personal website. Originally, 95.675 pages were intended to be included in Facebook page analysis. But after selecting category related page, only 2.793 pages are selected as candidate. From 3.314 pages and 991 users, a user-like matrix was generated.

4.4. Dimension reduction in user-like matrix

When we obtained a reduced user-category matrix, we restored the matrix into a user-Like matrix. Table 3 shows eight examples of users Like page from both political parties. User

![Figure 6. Importance to political attitude of Facebook page categories.](image)
from both parties has different pattern of likeness. For instance ‘Lee Jaesu-Citizen is the answer’ page is liked by Saejungchi supporters only. We also checked correlation among pages in order to avoid multi-collinearity problem.

After filtering out highly correlated pages, singular value decomposition was applied to reduce the dimensionality and reshape the page information into criteria. Suppose user-Like matrix $I$ be a $u \times p$ matrix as denoted in formula 2. Singular value decomposition results are three representation matrices,

$$L^C = U \Sigma V^T.$$  

Singular value decomposition dimension reduction quantization (Rajaraman & Ullman, Mining of Massive Datasets, 2002) set the values of matrix $\Sigma$ to zeroes in the corresponding rows and columns in the small singular values in $\Sigma$. The number of reduced singular values can be obtained by comparing the reduced matrix and original $M$ matrix. If the difference between original matrix and generated matrix is less than a chosen threshold, the reduced number of singular values of matrix $\Sigma$ can be used. Figure 7 shows how many pages that combined as criteria should be kept for 90% power of user-Like matrix representation.

4.5. Logistic regression and estimation

With remaining 284 criteria after SVD reduction, the stepwise model generation by minimizing Akaike information criterion (AIC) (Reddy & Reddy, 2010) is used for logistic regression and 10-folds cross validation to validate the model also AUC (area under the curve) to evaluate the model. The AIC optimization in stepwise logistic regression remains only 28 criteria (reliability $p < .05$). Figure 8 shows the AUC graph for the final estimation performance. We obtained AUC = 0.82, with average sensitivity 0.86 and average specificity 0.79.

The result of logistic regression is criteria with coefficient from each criterion towards political attitude (either supporter of ruling or opposite party). Coefficients larger than 1 are supporting parameters for ruling party, and coefficients less than 1 are considered for the opposition. Even though every criteria used in logistic regression are composed of multiple Like pages, we can deduce the major contributing Like pages from the SVD’s $V$ matrix (we used 0.5 for its criteria). There are some pages that represent the Saenuri party supporters and some pages that represent Saejungchi party supporters. For instance,
the pages that support Saenuri party are ‘Pop music’, ‘ZIEN ART SPACE’, ‘Pig-year born woman’, ‘GMA network’, and so on. In contrast, Saejungchi party supporting pages are ‘KNN’, ‘God’s Not Dead’, ‘Jetsetter’, ‘The Guy coming’, ‘Uppy Look’, ‘bed scene, how deep did you see?’, and so on in order of significance (Table 5).

Category-based space reduction step in the proposed algorithm reduces the computation significantly. All data analysis in this work is done using R on typical PC environment with Intel Core i7 CPU, 8GB RAM, 1TB SSHD and Windows 7 OS. In comparison with the direct SVD methods in (Kosinski et al., 2013), which failed finishing computation with the same computing resource, the full computation procedure is performed in less than 10 minutes. Furthermore, the two step approach provides the category significance information, which helps the general tendency more easily.

5. Conclusion and discussion

The contribution of the paper is two folds: developing a customized and computationally efficient political attitude estimation procedure on Facebook service, and examining Asian social-politics relations by applying the developed algorithm.

First, despite the emergence of big data computing platforms, such as map-reduce (Hadoop), computation algorithm study is still essential for utilizing non-massive data...
tools, like R language and MATLAB™. In this paper, we developed alternative solution to the problem in data collection and computational difficulties. Also the computation reduction algorithm provides important insight into meaning of the data set. The paper proposed a customized yet computationally light-weight procedure and algorithm for estimating attitude of the users through their Facebook Likes. After excluding likeness of political pages, the proposed algorithm obtained quite high estimation accuracy of AUC equal to 0.82. The authors believe that the same process is not restricted to the political attitude estimation, but can be applied to diverse domains, such as in comparing Samsung vs Apple, Brazil vs Argentina soccer team, Football vs Baseball.

Increasing use and study on online communications including SNS for politics are seen over the years, but a very few studies have been reported on the context of Asian countries. This study is focused on the data from one country only and the result may differ in data from other country because most of the user likeness that supports the political attitude. In this paper, dataset is related to local content and limited to the fact of Korea having two major political parties similar to the United States. However there are some countries that have multiple parties with the number of supporter distributed evenly as in Indonesia; hence we will require different analysis method unlike the logistic regression applied in this paper. Generalizing the analysis method that can be used flexibly in different type of political systems of different countries can be further improvement in this area of study.

However, it is not clear whether such a research strategy is useful in explaining public attitudes due to the limitation of available data. In other words, this study provides very limited information of political attitudes that this study aimed. Therefore, this study should be understood in methodological proposal rather than theoretical results.

Acknowledgements

H. Ahn designed the study. M. E. Wijaya developed system and data analysis. M. S. Billah joined the discussion and data analysis.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Muhammad Eka Wijaya has BS in information science from ITS, Surabay, Indonesia and MS degree from Seoul National University of Science and Technology in 2015.

Meer Sadeq Billah got MS from Seoul National University of Science and Technology in 2017 and he is studying for PhD in the same University.

Heejuone Ahn got all his BS, MS, PhD in Electrical Engineering from KAIST, and worked for LG Electronics and Tmax Soft Inc. Since 2004, he has been working as professor in Seoul National University of Science and Technology.

ORCID

Heejuone Ahn http://orcid.org/0000-0003-1271-9998
References


Kim, Y. (2011). The contribution of social network sites to exposure to political difference: The relationships among SNSs, online political messaging, and exposure to cross-cutting perspectives. Computers in Human Behavior, 27(2), 971–977.


Macafee, T. (2013). Some of these things are not like the others: Examining motivations and political predispositions among political Facebook activity. Computers in Human Behavior, 29(6), 2766–2775.


