PAPER *Special Section on Data Engineering and Information Management*

Culture Based Preference for the Information Feeding Mechanism in Online Social Networks

Arunee RATIKAN[†a)], ***Nonmember and*** **Mikifumi SHIKIDA**[††b)], ***Member***

SUMMARY Online Social Networks (OSNs) have recently been playing an important role in communication. From the audience aspect, they enable audiences to get unlimited information via the information feeding mechanism (IFM), which is an important part of the OSNs. The audience relies on the quantity and quality of the information served by it. We found that existing IFMs can result in two problems: information overload and cultural ignorance. In this paper, we propose a new type of IFM that solves these problems. The advantage of our proposed IFM is that it can filter irrelevant information with consideration of audiences' culture by using the Naïve Bayes (NB) algorithm together with features and factors. It then dynamically serves interesting and important information based on the current situation and preference of the audience. This mechanism helps the audience to reduce the time spent in finding interesting information. It can be applied to other cultures, societies and businesses. In the near future, the audience will be provided with excellent, and less annoying, communication. Through our studies, we have found that our proposed IFM is most appropriate for Thai and some groups of Japanese audiences under the consideration of audiences' culture.
key words: *Online Social Networks, information feeding mechanism, information overload, culture differences*

1. Introduction

Online Social Networks (OSNs), Facebook, Google+, Twitter, have recently become important communication media. They bring us closer together by enabling users to consume and share unlimited information. The OSNs are viewed as post owners site (creating and broadcasting the post to the OSNs) and audience site (receiving the posts from the owner of the post site). This research focuses on the audience site. IFM is an important part of the OSNs. Its main function is to serve information to the audience. The amount and quality of information depends on it. The IFM often feeds excessive information which may be unrelated to the actual needs of the audience. We believe the existing IFMs cause two problems in the OSNs: information overload and cultural ignorance.

Information overload [1], [2] is caused by sharing a huge amount of information, growth of social graph and not recognizing the audience's current situation and preference of IFM. Thus, audience's social network page (SNP) contains excessive information. Useless information causes annoyance, and leads to useful information missing. As a side-effect, audience activity is reduced [1], [3] in the OSNs. However, to reduce information overload in OSNs, we need to provide a sort of mechanism to filter out useless information. A suitable IFM should filter useless information by considering the real needs of the audience, and when and what kind of information the audience wants to read.

Up to present, cultural ignorance has not been emphasized by most existing IFMs when they serve the information to the audience. Audiences in different countries use their own cultural behavior and criteria when receiving the information. Many factors indicate cultural differences, i.e. age, career, etc. Reducing information by considering audiences' culture can alleviate information overload where an audience receives too much unnecessary information. In other words, we have to construct a IFM for the audience based on his/her culture and there is no universal IFM for all cultures. Each culture should be treated individually.

Existing works [2], [4], [5] attempt to solve the information overload problem by using filtering and recommendation. However, they do not realize the needs of the audience, i.e. current situation, preference, cultural differences, etc. Hence, they may not be efficient in reducing the huge amount of information, and serving relevant information to the audience. This research aims to develop a new type of IFM that attempts to solve information overload with consideration of culture dependency. We propose a concept that the information should be dynamically fed into the audience's SNP based on the current situation and ordered by audience's preference. Towards this, we have applied the NB algorithm together with different sets of influential features and factors to filter the information and evaluate our approach using subjects (audiences) of Japan and Thailand. Our proposed IFM can be adapted for a marketing and recommendation system. This paper is organized as follows: Section 2 shows the related studies. Section 3 presents our proposed method and details of the system architecture. Section 4 explains our experiments. Section 5 shows the results and measures the performance. Section 6 discusses the results. Finally, Sect. 7 concludes our research.

Manuscript received July 4, 2013.
Manuscript revised October 29, 2013.
[†]The author is with the School of Information Science, Japan Advanced Institute of Science and Technology (JAIST), Nomi-shi, 923–1211 Japan.
[††]The author is with the Research Center for Advanced Computing Infrastructure, Japan Advanced Institute of Science and Technology (JAIST), Nomi-shi, 923–1211 Japan.
a) E-mail: a.ratikan@jaist.ac.jp
b) E-mail: shikida@jaist.ac.jp
DOI: 10.1587/transinf.E97.D.705

2. Background and Related Works

2.1 Information Feeding Algorithms

There are many IFMs used for feeding information to the audience. Facebook uses EdgeRank algorithm [6]. It optimizes the audience's news feed by using scoring to make the decision as to what information should appear. It considers that item (photo, text, etc.) displayed on the audience's news feed is an object. When the object is interacted by other users such as commenting, tagging photo, etc., EdgeRank algorithm creates an edge. The edge composes of three components: affinity, weight and time decay. Affinity represents the audience's relationship with owner of item. The affinity is assigned by score. When the audience often connects to the owner of item, the audience will have a higher affinity score. Meanwhile, if the audience have not interacted with his/her friend for 1-2 year, the affinity score is very low. Weight means the audience's action affects weight. Interaction with video, photo and link is calculated as the highest weight. Time decay indicates how recent the item is. New item has more chance to appear on the news feed.

Google+ [7] allows the audience to see the posts from members in any circle via his/her Home Stream. Home Stream shows the posts, which come from specific or public sharing, according to time. However, the audience can filter the posts from selecting specific circles. Recently, the audience can control the number of posts appeared on the Home Stream by adjusting the volume (more, standard and fewer) in each circle, what's hot and communities. This indicates that Google+ no longer shows all posts in the audience's Home Stream.

Twitter [8], [9] uses a concept of social networking and SMS messaging. It indicates a simple way to provide others' status via Twitter timeline where it shows Tweets by using reverse chronological order. This means a new Tweet is added to its front. The audience can see the mix of Tweets in the timeline from following the people. Although Twitter allows the audience to see all of Tweets on his/her timeline, it provides search engine to filter out unwanted or unrelated Tweets and discover new Tweets.

2.2 Information Filtering Techniques

The concept of filtering reduces redundant and unwanted information before it is displayed to the user. Generally, three main techniques are used for information filtering. The *Content-based technique* uses the user's profile and content of an item to find regularities. This technique filters the information by using a data mining algorithm with features [2], [4], [5]. The *collaborative filtering technique* [10], [11] is usually adopted in OSNs such as personalized recommendation, Twitter user recommendation. The concept recommends the unknown item to the user by using user-item data [12]. The *hybrid collaborative filtering technique* is a combination of the first and second techniques. This technique [12], [13] can solve problems that the collaborative filtering technique cannot, and accurately recommends the item when a new item and new user just come on to the system. However, three techniques encounter the cold start problem, when no user has rated the item. Consequently, the system has difficulty matching the information to the item.

2.3 Cross-Cultural Differences

Hofstede [14] explains that people living in the same social environment may have similar cultural thoughts. Therefore, understanding regularity in such thoughts, feelings and actions of people in each culture is a challenge. Cross-culture research has been widely studied in education, m-commerce and so on. Studying how user culture affects usage of OSNs is another challenge task. For privacy concerns [15], Hong Kong users are more likely to disclose personal information to others, while France users feel less in control when updating personal information. A cultural effect on true commitment [16] means that US and UK users give priority to groups, while Italian users rate groups, games and applications as being the most important. As for the size of networks [17], culture has a great influence on relationship maintenance. The US students make less effort in taking care of their relationships, whereas the Korean students tend to get social support from existing relationships.

3. Proposed Method

Figure 1 displays an overview of system architecture for our proposed IFM. In this system, dynamically feeding interesting and important information has been performed by considering the audience's current situation and preference. Our proposed IFM consists of data mining, data collection, social graph generation, audience information, feature extraction and post organization. The core of the system is a data mining component. This component obtains useful information by data collection, social graph generation, audience information and feature extraction. After that, a post orga-

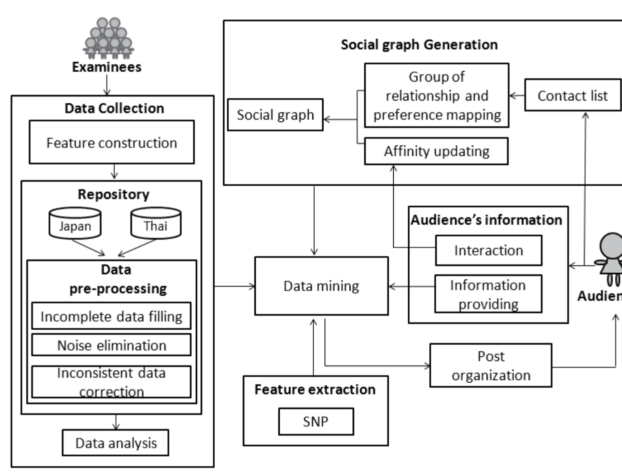

Fig. 1 High-level system architecture for proposed IFM.

nization component orders the posts which are allowed to show on the SNP according to audience's preference. Further, explanation of the six components is given below.

3.1 Data Collection

Its goal is to prepare a set of influential features and factors and study cultural differences between Japanese and Thai audiences in the OSNs. After analysis, all of the data is constructed as *training data* for the data mining component.

3.1.1 Feature Construction

In our work, we define and utilize seven types of features for filtering the information [3], [18]. However, these are still investigated in order to find influential features of the decision by the audience as to whether or not the post should be fed into their SNP. We also simulate several scenarios to indicate the characteristics of posts in the OSNs by those seven features as described in the first questionnaire in Sect. 4.

- **Audience's current situation (n_0)** indicates the activity being performed, i.e. work, private time, meetings, etc. Different situations influence the type of post the audience wants to read.
- **Topic of post (n_1)** refers to the post's category i.e. sport, music, advertisements, games, etc. Matching the topic of the post to preference will increase interest in reading the post.
- **Type of relationship between audience and owner of post (n_2)** is reflected in its association, i.e. boss, friend, family, etc. The audience makes a decision to read a post by considering relationship. This differs from existing OSNs where every member is a friend.
- **Affinity between audience and owner of post (n_3)** [6] shows the familiarity between the audience and owner of the post or how often they interact.
- **Affinity between audience and a person commenting (n_4)** can be considered when the post is created by the same owner of that post and gets similar popular level but different groups of commentators.
- **Post's Popularity (n_5)** refers to top stories currently of interest to many people. This feature relies not only on the number of comments or "like" button pressing, but also commentator interaction [6].
- **Time Decay (n_6)** [6] shows how up to date the post is. Recent posts tend to interest due to their newness.

3.1.2 Data Pre-Processing

The data collected from the first questionnaire, as explained in Sect. 4, is kept in separate databases for Japanese and Thais. In a real-word database, we need to ensure that our data is ready to use. To improve data quality, we need to normalize raw data, supply missing data, eliminate noisy data and correct inconsistent data.

3.1.3 Data Analysis

We use the feature selection tool WEKA [19] to investigate the hypothesis in Sect. 3.1.1. Similarities and differences between the Japanese and Thai cultures are then compared. Finally, these results are used as *training data*.

Table 1 indicates the first three features and factors [18] that are influential in examinees' decision making, and whether or not they allow certain kinds of posts to be fed into their SNP. The most important feature of both countries is clearly different in that the n_0 and n_1 features are given precedence by Japanese and Thai examinees respectively. The No.Times and age factors have a greater impact on examinees' decisions. We analyze which kind of posts causes annoyance to the examinees or should not be shown on the SNP when the examinees are in different situations by using the n_0 and n_1 features. We found that over 86% of Japanese examinees do not want to read advertisements about discount prices, part-time jobs and games during a meeting, at work or traveling. Clearly, around 90% of Thai examinees are not interested in advertisements about part-time jobs during meetings, work, private time or traveling. Also, about 80% of them ignore games during work time.

Figure 2 shows how the No.Times factor influences the examinees' decision. Graph patterns between both countries are different. When the No.Times gradually increases, Japanese examinees seem to answer "Allow". It is reasonable. Meanwhile, Thai examinees tend not to allow, especially at 10 times, since they are concerned about privacy and do not want to reveal their preference. Viewing this kind of post excessively on the SNP gives rise to boredom.

The results expose when the age increases, Japanese and Thai examinees pretty much answer "Not Allow", especially Thai examinees with an age higher than 30 years

Table 1 Influential features and factors in Japanese and Thai examinee's decisions.

Nationality	Features	Factors
Japanese	n_0, n_1, and n_6	The number of times a post is reviewed by the examinee (No.Times), age and preference
Thai	n_1, n_0, and n_2	No.Times, age and career

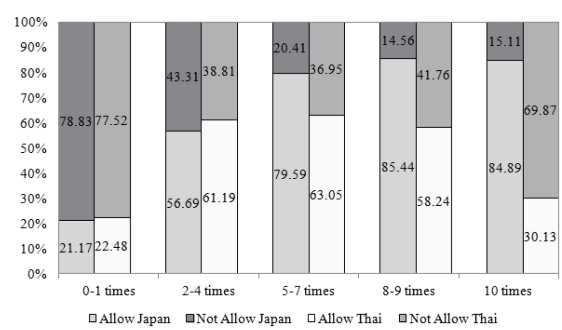

Fig. 2 The impact of the No.Times factor on Japanese and Thai examinees' decision making.

old. There is a relationship between age and preference from the interviews. In Thailand, young people have several interests. They are open to reading different things from the SNP. However, older people have specific interests and point to their interest during playing the OSNs. As for Japanese, most of them have their own specific interests.

For the n_6 feature, most of the Japanese examinees allow the posts which show as just having been created on the SNP because they like to know the latest information as quickly as possible according to the time series. In Japanese culture, punctuality is important in social aspects, i.e. transportation, business, etc., whereas Thai culture views time flexibility as being acceptable in certain situations.

For the career factor, we found the n_0 and n_2 features to be important in careers when selecting a post to read. In both countries, engineers and IT examinees have respectability at work. They restrict receiving posts during meetings and at work, especially from family [3]. Japanese culture in a workplace is taken more seriously than in Thai culture. It corresponds to Hofstede's individualism dimension [14]. It explains that in Japan, skill and performance are criteria in task assignment. Promotion relies on the seniority rule which recognizes age [20]. Japanese have a famous loyalty to the company, so they seldom change jobs or company. For Thais, a task is assigned to a group. Position promotion depends on performance and work period as important. This leads to competition for new positions or careers.

3.2 Social Graph Generation

Usually, users in existing OSNs are represented in several structures e.g. formal methods, matrices and graphs. In this research, we try to create a graph to represent the social relationship between the audience and members in a contact list. A node indicates each member. Any two nodes are linked with relationship in terms of affinity. This affinity is updated by audience interaction. This research also adds preference and type of relationship, e.g. boss, or friend to the graph, since these have to be used in the data mining component.

3.3 Audience Information

Personal information is required from the audience for the data mining component. For Japanese, the audience is asked about the n_0 feature and the No.Times, age, and preference factors. Nonetheless, Thai audiences also need to provide the n_0 and n_2 features and the No.Times, age, and career factors. The audience completes the age, career, No.Times and n_2 information only once, then the system remembers and uses it next time. In reality, the n_0 feature can be found automatically in the Google calendar or the calendar on the PC of the audience member. Moreover, the system observes audience behavior, and interaction with the post, to update the affinity aspect in the social graph.

3.4 Feature Extraction

Generally, any post created contains a lot of properties [21], i.e. post ID, time created, the name of the owner of the post etc. Some properties are not necessary for filtering the information, and thus we extract a lot of properties into meaningful features: the n_1 and n_6. For the n_1 feature, the post is classified into its topic based on the training data. Automatically, there are several possible algorithms used for text classification, and this is called supervised learning, i.e. Naïve Bayesian classification, Support Vector Machines (SVM), etc. Otherwise, we can categorize the topic of post by individual. For instance, if the post is "Shinji Kagawa expresses Sir Alex Ferguson's regret". This post will be classified as a "Sport" topic. For the n_6 feature, this can be directly extracted from the post, which is the created time.

3.5 Data Mining

This component is considered to be the core of our proposed IFM. It is important for automatic new and useful knowledge prediction from databases. Information filtering in this research uses the classification concept. Three algorithms are compared [19]: NB, Decision tree (DT), and K-Nearest Neighbor (KNN). The NB algorithm is based on conditional probabilities by applying Bayes' Theorem. Its advantages are fast, highly scalable model building, and reliable classification performance. The DT algorithm is a tree structure, and is the combination of mathematical and computational techniques. It has a high predictive performance and can handle a variety of input data. The KNN algorithm is a simple predictive algorithm using an entire training database as the model and decides classification using Euclid distance. It can predict discrete attributes and continuous attributes.

We use the classification tool of WEKA [19] to measure the performance of three algorithms. From different sets of features and factors for Japanese and Thais in Table 1, we add each feature and factor into three algorithms and observe the accuracy improvement of combinations when the number of features and factors changes. We have three combination groups: combinations of pure features and one factor, combinations of pure factors and one feature and combinations of features and factors. For combinations of pure features and one factor, the results show that if we use pure features (n_0 and n_1), they cannot give a high performance with accuracy 61.74% and 67.41% on average for Japanese and Thais respectively. However, a combination of pure features and one factor provides higher classification accuracy. When we add the No.Times factor into a combination of n_0 and n_1 features, of n_0, n_1 and n_6 features and of n_0, n_1 and n_2 features, the accuracy clearly increases by approximately 5–10%. It corresponds to the result in Table 1 that this factor has the most influence on audience decisions. Age, career and preference factors do not increase much.

For combinations of pure factors and one feature, the performance of combination of pure factors is higher than

Table 2 Accuracy of classification for three algorithm when considering combinations of the factors and one feature.

Combinations of the features and factors	NB		DT		KNN	
	JP	TH	JP	TH	JP	TH
1. n_0/n_1/No.Times/Age	75.75%	73.21 %	75.00 %	72.68%	75.7 %	71.01 %
2. n_0/n_1/No.Times/Age/Career	-	74.32%	-	73.31%	-	74.3%
3. n_0/n_1/No.Times/Age/Preference	76.45%	-	80.02%	-	79.67%	
4. n_0/n_1/n_6/No.Times/Age	77.95%	-	77.58%	-	77.81%	-
5. n_0/n_1/n_6/No.Times/Age/Preference	**78.53%**	-	79.71%	-	79.80%	-
6. n_0/n_1/n_2/No.Times/Age	-	71.92%	-	72.91%	-	71.01%
7. n_0/n_1/n_2/No.Times/Age/Career	-	**75.96%**	-	73.68%	-	73.68%

that of the combination of pure features with 75.84% and 71.41% on average for Japanese and Thais. We found that the number of factors affects the accuracy improvement for Japanese and Thais and three factors give a better performance than two factors. The feature extension has little influence on the accuracy for Japanese. For Thais, adding one feature into pure factors makes little improvement to the performance. We notice that increasing the n_1 feature into the No.Times, age and career factors gives a better result than adding the n_0 or n_2 features. This is because the n_1 feature has the highest impact on audience decision in Table 1.

For combinations of features and factors, we notice that they can obtain higher classification scores than the two previous combinations. As shown in Table 2, increase in accuracy comes from the number of combined features and factors. For example, when we add the preference factor into the combination 1 and 4 for Japanese, the classification accuracy is higher.

Based on the results of three groups, the NB algorithm obtains the highest classification accuracy with 72.32% and 71.11% on average for Japanese and Thais, while the KNN algorithm has the lowest classification accuracy. In terms of speed, the NB algorithm has the fastest speed for building the model, whereas the DT algorithm takes a long time; 28 times longer than the NB algorithm. We select the NB algorithm as having the best performance in classification for information filtering. For Japanese, we use the set of n_0, n_1 and n_6 features and the No.Times, age and preference factors although its accuracy, which achieves 78.53% is lower than the DT algorithm. For Thais, the set of n_0, n_1 and n_2 features and the No.Times, age and career factors gives the greatest accuracy at 75.96%.

3.6 Post Organization

Reverse chronological order and top stories are two of the main post organizations in existing OSNs. Reverse chronological order shows the post according to the time it is created such as Facebook, Twitter and Google+. The newest post will be presented at the top of the SNP. This means the audience always consumes the latest post. Usually, Facebook with this post organization displays updated posts from 250 friends and Facebook Pages. Top stories in Facebook show popular posts that are of interest to favorite friends. This post organization relies on factors such as the number of comments, as used in Facebook's EdgeRank algorithm [6]. Nonetheless, we propose any posts should be dynamically ordered by audience's preference. This is because the audience's preference can change all through the time. Moreover, in a real situation, reading an enormous amount of posts in the SNP is compared to finding the interesting posts, and depends on each audience's preference [18]. Therefore, this component allows the audience to set three preferences in descending sort. Then, it considers the topic of post, which is extracted in Sect. 3.4, when it orders the posts. For the posts, which do not relate to the audience's preferences, they will be sorted in reverse chronological order. This post organization leads to faster audience consumption of a large quantity of posts and is consistent with the requirements of the audience.

4. Experiments and Questionnaires

In our experiments, we use two questionnaires. Examinees in our experiment have good experience of Facebook, Twitter, Google+ and Mixi (Only Japanese examinees). They are students, engineers, and IT specialists.

In the *first questionnaire*, the objectives are to study and compare the cultures of Japanese and Thai examinees by using seven features as explained in Sect. 3.1.1. 51 Japanese and 110 Thai examinees are asked about demographic factors, i.e. age, career, interests, etc. Most of the Japanese examinees are aged between 23 and 29 years old and most of the Thai examinees are aged between 26 and 30 years old. Each examinee answers 45 questions that the examinees need to imagine themselves in the particular scenarios. The scenarios are randomly selected to the examinee. There are many types of scenarios, which come from combination of seven features, in this questionnaire. Each type of scenarios has an equal chance to be selected. We attempt to simulate each of them as much as possible to the examinee. Then the examinee answers the question: "If you see a post with different scenarios 10 times in the SNP, how many times do you review this post?" We also supply multiple choices to indicate the number of times. An example scenario is shown below.

"*There is a post about a* ***sport (team e.g. football, volleyball)*** *fed into your SNP. The owner of this post is* ***a friend from university*** *and you have* ***usually interacted with your friend every week or every month***. ***10-30 people comment or press a "like" button*** *on this post. You and another person commenting on this post have* ***regularly interacted every***

day. *This post has* ***just been created***".

Next, each examinee is asked "Will you allow this post to be fed into your SNP when you are **meeting with 10 co-workers**". The bold text means seven features. Then, the examinee makes a decision either to "Allow" or "NOT Allow". This scenario requires the examinee to imagine 10 times. This means that the examinee will face this kind of scenario with a different content but still on the same topic. For example, the topic of post is a sports team. The examinee can imagine that the post might be football or any other kind of sports team; however the main topic of the post remains the sport.

We have used seven topics in the experiment because they commonly occur in the OSNs.

- **Personal story** is described about personal experience, updating status, self-promoting, etc.
- **Advertisement** shows that recently the OSNs are used by companies, members in a contact list for online business, part-time job, or promoting their products or services by discount price. For example, Facebook allows the advertisements to appear because it believes [22] that "Everyone wants to know what their friends like".
- **Game** is kind of posting about game invitation such as Dragon city, FarmVille, etc. or opinion about game.
- **Working** is explained about task, schedule plan, etc.
- **Travel** presents a journey to some places where they are a famous place, an unseen place, a natural attraction, etc. Some of owners try to review the place where they have ever gone.
- **Sport** is kind of posting about football, basketball, volleyball, etc. Usually, most of people give feedback or opinion to those sports after they watched.
- **Music** refers to sharing or posting an interesting music, a popular music, an oldie music, etc.

The goal of the *second questionnaire* is to evaluate the performance of four IFM methods as shown in Table 3. This questionnaire is answered by 14 Japanese and 21 Thai examinees respectively. Each examinee is asked with 25 questions about demographic problems in current OSNs, and the performance of four IFM methods. Japanese and Thai examinees use OSNs for a specific purpose, such as entertainment, information sharing, consuming, business, for killing time and relationship maintenance. In addition, Japanese examinees use the OSNs to recruit new members. Most of the Thai examinees have more than 200 members in their contact list, while Japanese examinees have 20-100 members. We simulate the audience site in OSNs by showing the posts fed by each method. Method 1 is a nominal method with no exact specification. The information will be selected and shown to the examinee's SNP randomly; hence the examinee cannot expect information characteristic and ordering. The Method 2 applies timeline and reverse chronological order technique to show the information. The Method 3 uses EdgeRank algorithm for information feeding and top stories for post organization. The more details for Method 2 and 3 can be found in Sects. 2.1 and 3.6. Method 4 feeds the information to the examinees by using the NB algorithm together with different sets of features and factors for Japanese and Thais, and the personal information completed by the examinee as described in Sect. 3.3. This method uses the training data as mentioned in Sect. 3.1 and the test data from retrieving the examinee's data in Facebook. When the test data comes, it is identified to indicate the required features, which is necessary for classification. For example, the test data has to be categorized into its topic based on the training data. Then, it is classified by the NB algorithm. The Method 4 dynamically serves the information to the examinee, and this information is then ordered by the preference of the examinee. Furthermore, it allows the examinee to change their current situation and preference in relation to their requirements. For example, the NB algorithm classifies any two posts in Table 4 based on the information provided by the examinee: meeting (n_0), 2-4 times (the No.Times), 25 (age), engineer (career). The P1 will be shown on the examinee's SNP because of its high probability based on Bayes' Theorem [19].

Table 3 Description of information feeding and post organizations.

Method	Information Feeding	Post Organization
1	Random	Random
2	Timeline	Reverse chronological order
3	EdgeRank	Top stories
4	Our proposed IFM	Preference

Table 4 Example of classification result.

Post description	n_1	n_2	Result
P1.TMR has meeting at 1 pm	Working	Boss	Allow
P2.OMG!! I missed final train	Personal story	Friend	NOT Allow

To prevent bias, four name methods are not revealed. We use real data from each examinee via Graph API in Facebook developer [21]. The advantage of using real data is that the examinees do not need to imagine and are not to carry out the experiment due to data familiarity. However, we need permission from examinees due to privacy concerns.

5. Results and Performance Evaluation

Each question in the second questionnaire measures the ability of four IFM methods in Table 3 by Mean and Standard deviation. All questions are answered using Yes/No and a 5-point Likert scale (1=0% and 5=75-100%).

5.1 Opinion of the Problems in Current OSNs

Table 5 shows the examinees are encountering information overload on their SNP, and this information is not interesting to them. 64.29% and 71.43% of Japanese and Thai examinees express their annoyance at having their privacy disturbed. They miss useful information arround 4-6 times a day. Interestingly, 66.67% and 85.71% of Japanese and Thai examinees believe that feeding information to the SNP

Table 6 Performance evaluation by using mean and standatd deviation.

Evaluator	Method 1		Method 2		Method 3		Method 4	
	JP	TH	JP	TH	JP	TH	JP	TH
Information overload solution	2.43±1.22	2.43±1.12	**3.29±0.91**	2.62±1.02	2.43±0.85	3.19±0.93	2.64±0.93	**3.52±0.98**
Information filtering performance	2.50±1.22	2.57±1.12	2.79±0.97	2.67±1.11	3.00±1.17	3.10±1.22	**3.43±1.08**	**4.05±0.97**
Dynamical information feeding	1.86±0.86	2.67±1.28	3.36±0.93	2.95±1.12	2.86±1.10	3.14±1.06	**3.43±0.94**	**3.52±1.08**
Consistent information serving	2.93±1.14	2.81±1.33	**3.64±0.93**	3.19±0.98	3.43±0.94	4.04±0.80	3.57±1.02	**4.24±0.89**

Table 5 The examinees' opinion about information overload on the SNP.

Variables	JP	TH
Excessive information	3.36 ± 0.93	3.67 ± 1.11
Containing unwanted information	3.07 ± 1.00	3.47 ± 1.12
Privacy disturbance	64.29%	71.43%

based on their current situation is useful. Also, post organization has an impact on the reading content. Some Thai examinees said it orders the post according to priority, the most significant post is shown at the top of the page, whereas uninteresting posts are displayed at the bottom of the page.

5.2 Performance of Four IFM Methods

Table 6 reveals significant difference between Japan and Thai. Each question contained by evaluator is asked independently. For overall performances, the Method 1 is the worst performance for both countries because it does not use any algorithm for feeding the information. Other methods will be described relying on each evaluator below because each method has different advantages and depends on attitude of each culture or life style.

- **Information Overload Solution.** In this evaluation, we ask about the ability of each method to solve the information overload problem. For Japanese, after taking the ANOVA test, four methods have no significant differences, $F(3, 52) = 2.352$, $p > 0.05$. Most of them think Method 2 (3.29 ± 0.91) can solve the information overload problem. This is relevant to the previous analysis in Sect. 3.1.3. In further analysis, we found age, career and gender have no influence on the results. For Thais, there is at least one significant difference among the four methods, $F(3, 80) = 5.21$, $p < 0.05$. Scheffe values show that Method 1 (2.43 ± 1.12, $p = 0.01$) and Method 2 (2.62 ± 1.02, $p = 0.047$) are statistically significantly lower than Method 4 (3.52 ± 0.98). Most of the examinees think Method 4 is the most appropriate in solving excessive information feeding in Table 5 due to its flexibility. The quantity of information can be changed according to the current situation. When we analyze careers, the results reveal that most of the examinees aged between 26 and 30 years old or engineers, and believe in Method 3.
- **Information Filtering Performance.** The examinees are asked about the performance of each method in removing the uninteresting and unimportant information according to current situations or needs. The results of all four methods are not statistically significant in this evaluation for Japanese, $F(3, 52) = 1.707$, $p > 0.05$, while four methods show significant differences for Thais, $F(3, 80) = 7.755$, $p < 0.01$. A post hoc test indicates that Method 4 (4.05 ± 0.97) is significantly higher than Method 1 (2.57 ± 1.12, $p = 0.001$) and Method 2 (2.67 ± 1.11, $p = 0.002$). Method 3 is almost significantly different at $p = 0.06$ from Method 4. The overall performance clearly shows that Method 4 (3.43 ± 1.08 Japanese, 4.05 ± 0.97 Thais) overcomes the remaining methods. It filters uninteresting and unimportant information by using the NB algorithm. We have analyzed the examinees' current situations. The results show that most of Japanese and Thai examinees need information filtering especially when at work or in meetings. Beside the quality of information, the number of members in a contact list might create the need for information filtering. Presently, most Japanese use OSNs besides Mixi such as Facebook and Twitter, and getting new members is one of the purposes in using the OSNs. Most of the Thai examinees have more than 200 members on average in the contact list. Hence, they have a high chance of receiving excessive information and need to filter it.
- **Dynamically Information Feeding.** This evaluation measures how each method can dynamically serve the information to examinees based on their current situation or needs. Four methods for Japanese have statistically significant differences, $F(3, 52) = 7.952$, $p < 0.05$. Method 1 (1.86 ± 0.86) has significant differences with Method 2 (3.36 ± 0.93, $p = 0.002$) and Method 4 (3.43 ± 0.94, $p = 0.001$), whereas there are no statistical differences between the four method results in this evaluation for Thais, $F(3, 80) = 2.099$, $p > 0.05$. However, Method 4 (3.43 ± 0.94 Japan, 3.52 ± 1.08 Thais) shows the highest performance for both countries. The NB algorithm uses the examinee's current situation, which is one set of features and factors of classification, and therefore when the examinee's current situation changes, the information fed into the SNP also changes. Nevertheless, we cannot discard Method 2 for Japanese because its mean score is closer to Method 4. We found that most of the Japanese engineers and IT specialists or females said that when they open the OSNs, the information is dynamically fed according to time change. Consequently, they can consume updating information.
- **Consistent Information Serving.**
  Table 5 reveals that the examinees' SNP contains unwanted information, therefore the quality of information according to their requirements is important. For

Japanese, there are no statistically significant differences between the four methods, $F(3, 52) = 1.425$, $p > 0.05$. For Thais, the performance of the four methods has significant differences in this evaluation, $F(3, 80) = 9.397$, $p < 0.05$. By running post-hoc tests, Method 4 (4.24 ± 0.89) is statistically significantly higher than Method 1 (2.81 ± 1.33, $p = 0.0001$) and Method 2 (3.19 ± 0.98, $p = 0.015$). Also, Method 3 is significantly different to Method 1 ($p = 0.003$). Table 6 shows that most of the Japanese examinees believe Methods 2-4 have effectiveness in serving interesting and important information by using different concepts. Nevertheless, Method 2 possesses the highest mean score (3.64 ± 0.93). We found that career and age have an impact on the results. Japanese examinees, who are students or aged between 21 and 25 years old, think Method 4 has the best performance. For Thais, although Method 4 (4.24 ± 0.89) still satisfies the examinees, we cannot ignore Method 3 (4.04 ± 0.80), which is slightly lower in the mean score than Method 4. Most of the female examinees trust Method 3. From the interview, they are usually interested in entertainment or fashion, and therefore these can be out of date.

6. Discussion

From analysis and interview, Japanese and Thai examinees show significant differences in the selection of suitable IFMs. For Japanese, it is not clear which method is the most appropriate. However, we found that career and age have an influence on the overall performance in four evaluations. Examinees older than 26 years of age, engineers or IT specialists believe that Method 2 helps them to get the latest information, which they can follow in real-time. Especially during the working day, when they are very busy, they need to consume the entire information in a short message and in a short time. This indicates that time is important to them. Nonetheless, examinees aged between 21 and 25 years old, or students, like Method 4. They state that they can get suitable information about what their friends are doing in current situations automatically and dynamically. They do not want to select any information to read, but they need the IFM to choose it for them. This is because these groups of examinees usually have many different periods during the day such as a class, a seminar in a laboratory and so on.

Meanwhile, Method 4 clearly satisfies Thai examinees. It can solve the information overload problem because it can reduce the quantity of information on the SNP according to the current situation. Also, it filters inconsistent information, and then serves interesting and important information by using preference as post organization. Therefore, the examinees can dynamically receive the information based on their current situation and preference. This can be compared to reading a newspaper, where the information in the OSNs is generally diverse, i.e. news, events, entertainment, etc. The information served by Method 4 will bring the audience up to date without the need to read newspapers. Moreover, it helps the examinees to save time in finding interesting information and increases the opportunity to get new information casually. The benefits of Method 4 are suitable for the lifestyle of Thai examinees, as they usually try to adapt to various situations. Hence, the IFM should allow them to control information on their SNP independently. From analysis, career and gender impact slightly on the overall results.

Furthermore, we observed the examinees' behavior. When we asked them to carry out the experiment, almost all of the Japanese examinees asked "when is the deadline?" and answered "yes". This shows the importance of time and style of answer. Some of them are anxious if they are not sure whether or not they will finish the experiment in time. This behavior is explained in Hofstede's uncertainty avoidance dimension [14] as Japanese worry about unexpected circumstances. Hence, they try to prepare themselves for any eventuality. Although they are worried, they say "yes". In Japanese culture, negative words are considered impolite, so it is necessary to use clues to infer real meaning. Around 80% of Thai examinees state "when I have time, I will do it". Most of Thai people favor flexibility. They always make adjustments to suit a person or situation. Time or deadlines in Thailand are changeable [14]. For example, in business, parties may not be able to adhere to an exact deadline in negotiations [23] since Thai business people prefer long-term business arrangements to obtain long-term benefit.

Our proposed IFM can be applied to other cultures or societies with similar characteristics to Thais such as lifestyle, business negotiation, or working practices. People in Southeast Asia share cultural traits, social freedom and climate. For example, Thailand and Vietnam are similar in business negotiation. Our proposed IFM is also suitable for an audience with various daily activity periods such as engineers, or business people. Another benefit of our data analysis is the profitability potential to companies, provided entrepreneurs have a good marketing strategy. They need to understand the appropriate periods of time for optimum promotion of their products. For instance, advertisements should not be posted during working day. For recommendation, if a system knows the audiences' preference, it can suggest unknown information by the known preferences of an audience group, i.e. Amazon, eBay, etc.

7. Conclusion and Future Direction

We propose a new type of IFM in order to solve the information overload and cultural ignorance problems. Our proposed IFM filters uninteresting and unimportant information in order to reduce the amount of information on the SNP by the NB algorithm together with features and factors. The audience can dynamically consume the interesting and important information in a short space of time based on the current situation and have a better chance of obtaining new information casually. This information is ordered by audience's preference. Our proposed IFM can improve the limitation of existing IFMs such as information overload, lack of information, and providing uninteresting and unimportant information. It can also be applied to other cultures, societies and businesses. According to the results, our proposed IFM is the most appropriate for Thais and some groups of Japanese audiences. In future, we plan to combine our proposed IFM with other post organizations to evaluate its performance. Moreover, we plan to gather posts with the same idea into one post in order to save time in reading and thereby reducing the number of posts.

Acknowledgement

This research is supported by JAIST. We would like to express our gratitude to all the Japanese and Thai examinees who took part in the experiment.

References

[1] K.H. Koroleva, Ksenia, and O. Gunther, " "Stop me!"-exploring information overload on facebook," 16th Americas Conference on Information Systems, paper 447, 2010.
[2] K. Koroleva and A. Bolufe-Rohler, "Reducing information overload: Design and evaluation of filtering & ranking algorithms for social networking sites," 20th European Conference on Information Systems, paper 12, 2012.
[3] A. Ratikan and M. Shikida, "Feature selection based on audience's behavior for information filtering in online social networks," 7th International Conference on Knowledge, Information and Creativity Support Systems, pp.81–88, Australia, 2012.
[4] S. Nakamura and K. Tanaka, "Temporal filtering system to reduce the risk of spoiling a user's enjoyment," 12th International Conference on Intelligent User Interfaces, pp.345–348, Honolulu, 2007.
[5] S. Loeb and E. Panagos, "Information filtering and personalization: Context, serendipity and group profile effects," IEEE Consumer Communications and Networking Conference, CCNC, pp.393–398, Las vegas, 2011.
[6] Facebook EdgeRank and GraphRank Explained. http://blog.involver.com/2011/10/25/facebook-edgerank-and-graphrank-explained.
[7] GooglePlus:View and filter your Home page with circles. https://support.google.com/plus/answer/1269165?hl=en/
[8] J. Hannon, M. Bennett, and B. Smyth, "Recommending twitter users to follow using content and collaborative filtering approaches," 4th ACM conference on recommender systems, RecSys '10, pp.199–206, New York, USA, 2010.
[9] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?," Proceedings of the 19th International Conference on World Wide Web, WWW '10, pp.591–600, New York, NY, USA, ACM, 2010.
[10] A. Deng, Y. Zhu, and B. Shi, "A collaborative filtering recommendation algorithm based on item rating prediction," Journal of Software, vol.14, no.9, pp.1621–1628, 2003.
[11] C. Yang, J. Sun, and Z. Zhao, "Personalized recommendation based on collaborative filtering in social network," IEEE International Conference on Progress in Informatics and Computing (PIC), pp.670–673, 2010.
[12] X. Su and T.M. Khoshgoftaar, "A survey of collaborative filtering techniques," Adv. in Artif. Intell., vol.2009, 2009.
[13] R. Burke, "Hybrid recommender systems: Survey and experiments," User Modeling and User-Adapted Interaction, vol.12, no.4, pp.331–370, 2002.
[14] G. Hofstede, G.J. Hofstede, and M. Minkov, Cultures and Organizations: Software of the Mind: Intercultural Cooperation and Its Importance for Survival, McGraw-Hill, New York, 2010.
[15] H.K. Tsoi and L. Chen, "From privacy concern to uses of social network sites: A cultural comparison via user survey," PASSAT/SocialCom 2011, pp.457–464, Hong Kong, 2011.
[16] A. Vasalou, A.N. Joinson, and D. Courvoisier, "Cultural differences, experience with social networks and the nature of "true commitment" in facebook," Int. J. Hum.-Comput. Stud., vol.68, no.10, pp.719–728, 2010.
[17] Y. Kim, D. Sohn, and S.M. Choi, "Cultural difference in motivations for using social network sites: A comparative study of american and korean college students," Comput. Hum. Behav., vol.27, no.1, pp.365–372, 2011.
[18] A. Ratikan and M. Shikida, "A study of cross-culture for a suitable information feeding in online social networks," 15th International Conference on Human-Computer Interaction, USA, pp.458–467, Springer-Verlag, Berlin Heidelberg, 2013.
[19] I.H. Witten, E. Frank, and M.A. Hall, Data mining: Practical machine learning tools and techniques, Morgan Kaufmann, 2011.
[20] R.J. Davies and O. Ikeno, The Japanese mind: Understanding contemporary Japanese culture, Tuttle Pub, Boston; Tokyo, 2002.
[21] Facebook developer. http://developers.facebook.com/
[22] Facebook Ads. https://www.facebook.com/settings?tab=ads/
[23] L. Katz, Negotiating International Business, Booksurge, 2007.

Arunee Ratikan received B.S from Thammasart University in 2008. She received a scholarship from TAIST during 2008–2010. She is now a doctoral candidate at JAIST under the Graduate Research Program. Her research interests include Online Social Networks, HCI, cross-culture differences.

Mikifumi Shikida received his Ph.D in Engineering from Tokyo Institute of Technology in 1995. In 1995, he became a research associate at the Center for Information Science, JAIST. Since 2012, he has worked as a professor at the Research Center for Advanced Computing Infrastructure, JAIST. His research interests include groupware and large-scale network services.