PAPER Special Section on Log Data Usage Technology and Office Information Systems

Upcoming Mood Prediction Using Public Online Social Networks Data: Analysis over Cyber-Social-Physical Dimension

Chaima DHAHRI[†a)], *Nonmember*, **Kazunori MATSUMOTO**[†b)], *and* **Keiichiro HOASHI**[†c)], *Members*

SUMMARY Upcoming mood prediction plays an important role in different topics such as bipolar depression disorder in psychology and quality-of-life and recommendations on health-related quality of life research. The mood in this study is defined as the general emotional state of a user. In contrast to emotions which is more specific and varying within a day, the mood is described as having either a positive or negative valence [1]. We propose an autonomous system that predicts the *upcoming* user mood based on their online activities over cyber, social and physical spaces without using extra-devices and sensors. Recently, many researchers have relied on online social networks (OSNs) to detect user mood. However, all the existing works focused on inferring the current mood and only few works have focused on predicting the upcoming mood. For this reason, we define a new goal of predicting the upcoming mood. We, first, collected ground truth data during two months from 383 subjects. Then, we studied the correlation between extracted features and user's mood. Finally, we used these features to train two predictive systems: generalized and personalized. The results suggest a statistically significant correlation between tomorrow's mood and today's activities on OSNs, which can be used to develop a decent predictive system with an average accuracy of 70% and a recall of 75% for the correlated users. This performance was increased to an average accuracy of 79% and a recall of 80% for active users who have more than 30 days of history data. Moreover, we showed that, for non-active users, referring to a generalized system can be a solution to compensate the lack of data at the early stage of the system, but when enough data for each user is available, a personalized system is used to individually predict the upcoming mood.

key words: *mood prediction, pearson correlation, OSNs analysis, autonomous system*

1. Introduction

With the appearance of context-aware recommendation systems (CARS) [2], researchers have given more attention to user-context analysis in order to provide more accurate recommendations. In such systems, learning user preferences is a crucial step. With the growth of online social networks (OSNs) and the increasing number of active users, online activity behavior becomes a rich source of real-time data from which user preference can be learned. By posting on OSN accounts, users express their likes, attitude, activities, etc. resulting in a huge volume of data distributed across OSNs. Despite its abundance, this data has an obvious user-centric characteristic which make it more complicated to analyze.

Many works have been done for mood prediction. However, most of them have been focused on collecting data using extra devices like smartphones, body sensors, microphones and cameras [3], [4] or using linguistic features extracted from user's online posts to infer the current user's mood [5], [6]. Collecting such data is usually aggregated per day due to its sparseness or to its energy consumption. For this reason, previous methods can only predict the mood of today at the end of the day. Although recognizing the current mood is important in some domains, it does not provide any insight for other applications such as recommendation systems where the upcoming context (i.e. mood) matters to make customized recommendations for users. Furthermore, predicting tomorrow's mood given today's data have a number of important clinical benefits. This type of prediction provides an estimate of a person's future wellbeing which could potentially allow them to better prepare for it, or even better, change it. It is just as a weather forecast gives one a chance to take an umbrella rather than being left to be soaked by the rain [7]. Predicting short term mood developments among depressed patients is defined in the literature as the task to predict the value reported for the mood of the patient on the next day [8]. For the reasons mentioned above, we set the subject of prediction to tomorrow's mood instead of upcoming in a few hours.

Upcoming mood prediction plays an important role in diverse topics such as emotion variety impacts on behavior and decision making [9], bipolar depression disorder in psychology [8], and recommender systems in artificial intelligence [10]. Early detection could contribute to appropriate interventions by warning a user that a fall-back might be expected during the next days, or by applying a strategy to prevent critical situations [8]. For example, in quality-of-life recommendation systems, forecasting the mood is essential to develop actions to boost or to ameliorate the user's mood before its degradation. Based on what was described above, it is be worthwhile to forecast the user's mood instead of recognize the current one. To be independent from any devices, we opt for analyzing OSNs data to forecast user's mood. To do so, we expand the already existing set of features (linguistic-based features) by extracting three types of features from the user's online data.

In this paper, we verify the possibility to predict the upcoming mood of each user (positive, negative and neutral) by analyzing their *public* OSNs activities on different dimensions, namely cyber, social and physical spaces without using extra devices/sensors. The cyber space refers to the user online activities. The social space is related to the

Manuscript received November 12, 2018.
Manuscript revised March 30, 2019.
Manuscript publicized June 21, 2019.
[†]The authors are with the KDDI Research, Fujimino-shi, 356–8502 Japan.
a) E-mail: ch-dhahri@kddi-research.jp
b) E-mail: matsu@kddi-research.jp
c) E-mail: hoashi@kddi-research.jp
DOI: 10.1587/transinf.2018OFP0006

user social connectivity, such as people accompanying the user, etc. The physical space shows the activity and/or the location (restaurant, for example) in real world.

To validate our proposal, we have conducted an experiment over two months where we collected user's daily self-report about their mood and recorded their activities on OSNs. Next, we studied the correlation between tomorrow's mood (positive, negative and neutral) and today's online activities on OSNs for each individual user and for different groups, namely women, men, active users, non-active users. Finally, we built two predictive systems; personalized and generalized systems. The personalized system is trained for each user separately using their own public data. If a user does not have enough history data, we refer to the generalized system to infer their mood. It uses the data collected from a set of users to train the predictive system. The results showed that tomorrow's mood can be predicted using user's public online data over cyber, social and physical spaces. We showed, also, how to overcome the sparsity of data at the early stage of the system by using a generalized predictive system per gender.

To summarize, we proposed the extraction of cyber, social and physical features from user's OSNs accounts to train a generalized and a personalized mood predictive system that predicts the **upcoming mood** of a user. We compared the performance of our systems to the performance of two conventional linguistic-based methods. Our proposed method shows a better performance in terms of precision, recall and accuracy than the linguistic-based methods.

2. Related Work

Previous works can be categorized into two categories; current mood prediction and upcoming mood prediction.

2.1 Current Mood Prediction

Many works in the literature analyzed user's online data to predict its psychological state. These works focused on:

- Psychological state like personality traits of a user (OCEAN) [11], [12]. To do so, the authors in [11] studied the correlation between OCEAN and the type of Twitter user (Listener, Popular, Highly-read and influential). However, in [12], the correlation was calculated between OCEAN and some features extracted from Instagram images.
- Psychological illness detection like depressive disorder [13] or postpartum depression in new mothers [14]. The authors use features related to emotion, engagement, ego-network and linguistic style to classifier users into depressive and non- depressive classes [13] and to detect the change in emotions in new mothers [14].
- Current mood or change of mood prediction [5], [6], [15]. A collective sentiment study was considered in [15] where the authors compare sentiment change over multiple topics (iPhone, Android, Blackberry). M. Roshanaei explored a number of potential features related to mood [5]; linguistic, psychological features, gender, diurnal online activity and personal activity using Linguistic Inquiry and Word Count (LIWC) dictionary. No results were shown related to correlation or prediction. A more recent work was reported in [6] where the authors extracted features based on online activity on Twitter and Facebook to track the change of mood of a user. However, they collected a very specific dataset (32 highly active students) in a specific period of time (exam then vacation). The data was aggregated using a window of 7 days and results show that only 16 students were highly correlated with the extracted features.
- Mood prediction from phone usage data [3], [16]. In these works, an application should be installed on the user's mobile phone to collect phone usage data. In [16], the authors recorded the user's social interaction by counting the number of messages, emails, duration of calls. In order to predict user's mood, the application should access to the content of messages to count the number of words and characters used which may cause privacy issues.

In all papers mentioned above, the eventual goal was predicting the current mood and not the upcoming one. In addition, predicting user's mood from phone data requires the installation of a specific application which consumes more energy depending on the frequency of data collection and may cause some privacy issue when accessing to the user's private data. Moreover, OSN-based methods, mentioned above, assume the existence of text-based content in user's posts to extract linguistic features, which makes them not effective in cases where text-based content is not available. An example of such information is a check-in post on OSNs such as Instagram or Foursquare, without any textual comments from the user.

2.2 Upcoming Mood Prediction

Despite the large amount of works on mood prediction, forecasting the mood has not been well studied. Existing works focused on forecasting a health care problem such as forecasting severe depression based on self-reported histories [17]. Suhara et al. developed a smartphone application to collect user inputs about mood, medication records, sleeping hours, etc. The results show that two weeks history data are indicative to forecast depression. The method described in DeepMood [17] suffers from the amount of noise introduced by the user. Since users are inconsistent in giving their history, the effectiveness of such system is questionable. Besides, self-report poses additional burden for users which makes them unsatisfied with the system. More recent work [7] focused on predicting tomorrow's mood (happy/sad), stress (stressed/calm) and health (sick/healthy) by extracting a large number of features from smartphone logs, physiological sensors and behavioral surveys. This method is not practical since the user must wear and carry extra devices/sensors.

3. Proposed Method

The essential points of the proposed method are the extraction of features from multiple dimensions and the autonomy of the system:

- Extracting three types of features to the linguistic type of features: cyber, social and physical (location). Existing works focused on analyzing the content of a user post (usually tweets from Twitter) by extracting linguistic and psychological features and/or features related to cyber activities (count of positive/negative tweets, count of active/passive posts, time of post, etc.). However, users are using different types of social networks that generate data with multiple dimensions on cyber, social and physical spaces. Having multiple accounts, a user can post on one account or on several accounts, can post either text-only, image/link/hashtag-only or multi-modal (text and image/video) posts. Let's consider the scenario where the user posts a non-text tweet on Twitter or shares its location from Foursquare. Conventional works cannot predict the mood because, no text is available to analyze, in this case. However, in our method, we rely, not only on cyber features but also on features related to physical (check-in) and social (mentioned people) spaces. These features allow us to predict the mood even when the user is not tweeting that day by looking at his location or people he is with.
- Building an **autonomous method** that predicts the mood without the need of any extra devices and applications installed on user's mobile phone.

3.1 Methodology

In order to predict the upcoming mood of a user, we build a multi-stage system as shown in Fig. 1. First, for each user, we collect their IDs on Twitter and Instagram. We download their posts and activities on OSNs within the study period. We extract not only linguistic features but also cyber, social and physical types of features from each user's post. Then, we calculate the correlation (Pearson coefficient ($r \in [-1, 1]$)) between these features and the user's mood of the following day (see Fig. 2). The result of the correlation will be used to train the predictive system. Only significantly correlated features are selected and concatenated in a vector-like format, which is used to train the mood predictor. The predicted mood $\tilde{m}(u, t)$ for user u at day t is defined as:

$$\tilde{m}(u, t) = Predict(F(u, t-1)) \qquad (1)$$

where *Predict* is the learning algorithm used for prediction and F is the set of features defined as:

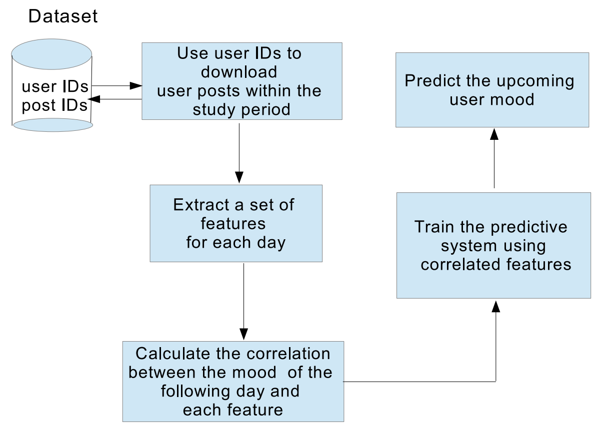


Fig. 1 Flow chart of the mood prediction system

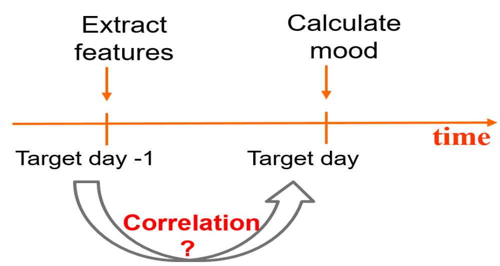


Fig. 2 Correlation between extracted features and target mood

$$F(u, t-1) = < f_1(u, t-1), f_2(u, t-1), \ldots, f_n(u, t-1) >;$$

n is the number of features and $f_i(u, t-1)$ is the feature number i, extracted at day $t-1$, for user u.

3.2 Online Data

Online data includes data about the participants' online activities on Twitter and Instagram (*public* data only). Since all participants are from Japan, the collected posts were in Japanese. Two crawlers were developed to collect all users' posts on a daily basis.

For Twitter, we used Twitter search API to collect all tweets posted during the study. For Instagram, we downloaded, during the study, all users' posts including images captions. The collected data was later aggregated per day for each user.

3.3 Feature Extraction

As claimed in our proposed method, we extracted not only linguistic but also non-linguistic features that can be used to predict the mood in days where non-text data is available. To forecast user's mood on a specific day, we extract features from OSNs activities of the previous day. We explored the following three types of features to design our predictive system. Cyber features were selected according to the literature (count of positive/negative tweets, count of active/passive posts, time of post, etc.) [6], [18]. The count of positive and negative posts are calculated by applying sentiment analysis to tweets and Instagram posts (caption) using

Table 1 Foursquare venue category list

ID	Venue Category	List of Venues
1	Arts & Entertainment	Amphitheater, Aquarium, Art Gallery, Circus, Concert Hall, Karaoke Box, Movie Theater, Historic site, Museum, Music venue, Pachinko, Public art (street art, sculpture), Stadium, Theme Park, Water Park, Zoo.
2	College & University	
3	Event	Christmas Market, Conference, Festival, Line/Queue, Sale, Parade, Street fair.
4	Food	Restaurant, Coffee shop, Bakery, Dessert shop, Ice cream shop, Donut shop, Diner, Food truck, Juice bar, Tea room, Snack
5	Nightlife Spot	Bar, Lounge, Night Market, Nightclub
6	Outdoors & Recreation	Athletics & Sports, Bathing Area, Bay, Beach, Bike Trail, gardens, Bridge, Cave, Castle, Farm, Forest, Fountain, Harbor / Marina, Hot Spring, Indoor Play Area, Playground, Light house, Lake, Pool, Rafting, States & Municipalities, Summer camp, Ski, Nature, Park, Waterfall, Well
7	Professional & Other Places	Auditorium, Building, Business Center, Meeting Room, Government Building (Town Hall, Police station), Industrial Estate, Library, Medical Center (Doctors, hospital), Parking, Post office, Power plant, Schools, Church, Mosque, Temple, Wedding Hall, Landmark.
8	Residence	Home (private), Residential Building (Apartment / Condo), Assisted Living.
9	Shop & Service	ATM, Adult Boutique, Chocolate shop, Women's store, Cosmetics, Arts & Crafts Store, Auto Garage, Bank, Bike Shop, Book store, Accessories Store, Convenience Store, Discount Store, Drugstore, IT Services, Lottery, Mobile Phone Shop, Nail Salon, Shopping Mall.
10	Travel & Transport	Airport, Baggage Locker, Bike Rental / Bike Share, Boat or Ferry, Bus Stop, Cruise, Hotel, Taxi, Train.

Table 2 Features description and statistics for Pearson's correlation between feature and tomorrow's mood

Feature	Social Network	Feature Name	Feature Description	Mean(r)	Variance(r)	standard error(r)
Cyber	Twitter	Negative text count	the count of negative tweets on Twitter per day	0.52	0.11	0.06
	Instagram	Negative caption count	the count of negative caption on Instagram per day	0.51	0.01	0.01
	Twitter	Positive text count	the count of positive tweets on Twitter per day	0.64	0.11	0.06
	Instagram	Positive caption count	the count of positive caption on Instagram per day	0.61	0.05	0.02
	Twitter	Hashtag count	the count of hashtags per day	0.56	0.07	0.05
	Twitter	Media count	the number of media attached to all tweets in a day	0.42	0.01	0.008
	Twitter	Source	the number of tweets, per day, generated from applications other than Twitter	0.53	0.05	0.01
	Twitter	Weekday v.s. weekend	1 if the day is weekday (Monday to Friday) and 0 if weekend (Saturday and Sunday)	0.55	0.08	0.03
	Twitter	Day v.s. night	the ratio of number of tweets at daytime (from 6 am to 8 pm) over nighttime (from 8 pm to 6 am)	0.56	0.09	0.03
	Twitter	Active v.s. passive	the ratio of active actions over the sum of active (posts done by user) and passive actions (retweets)	0.54	0.06	0.01
	Instagram	Hue	the average hue value over all images posted in a specific day (a value between 0 and 1).	0.59	0.06	0.03
	Instagram	Saturation	the average saturation value over all images posted in a specific day (a value between 0 and 1)	0.59	0.06	0.03
	Instagram	Brightness	the average brightness value over all images posted in a specific day (a value between 0 and 1)	0.69	0.04	0.02
Social	Twitter	Mention count	the number of mentions per day	0.47	0.09	0.04
	Twitter	Favourites count	the number of favourite per day	0.47	0.05	0.01
	Instagram	Like count	the number of like aggregated per day	0.58	0.07	0.03
	Instagram	Comment count	the number of comments on Instagram post aggregated per day	0.46	0.05	0.03
Physical	Instagram	Location tags	10-dimensional feature in which each dimension expresses a Foursquare venue category (1: visited, 0: non visited)	0.78	0.02	0.01
	Instagram	Geo-tag shared	1 if the user's post on Instagram has a geo-tag information and 0 otherwise	0.38	0.007	0.004

Google Cloud Natural Language API[†]. The social features are related to the user's social connectivity, such as people accompanying the user (mentions). The physical features show the activity and/or the location (restaurant, for example) in real world. The detailed list of the extracted features

[†]https://cloud.google.com/natural-language/

is described in Table 2. The list of the extracted features has more cyber features than other types of features, because social and physical features are more difficult to extract from public SNS data. To extract more social features (list of friends, like list of your friends, etc) and physical features (location data) we need to collect non-public information.

In this work, we asked three labelers to extract location data from Instagram posts. The labeling task corresponds to giving one or multiple labels to each post referring to Foursquare venue category list (see Table 1). We consider Foursquare to reflect people actual activities, because we think that the venue category such as food, university and professional places is highly correlated to high-level activities such as eating, studying and working. To do so, the labelers rely on the geo-tag information (if available), caption and hashtags associated with each post. To each post, we associate the labels that have a high labeling agreement from all labelers. This feature can be extracted automatically by using a machine learning algorithm to categorize user's location information into one of Foursquare venue categories. This, however, is beyond the scope of this paper. When a feature is not available, we append '0' or 'false' in the feature vector, depending on feature's definition. All the features in Table 2 are normalized using z-score normalization.

3.4 Prediction Algorithm

We tried different prediction algorithms including linear regression, k-NN, SVR, multi-layer perception, REPTree and Random Forest. Referring to the results presented in our previous work [19], we found that Random Forest was the best performing algorithm. For this reason, we chose Random Forest as prediction algorithm in this work.

4. Data Collection

4.1 Experiment Details

Our goal is to predict the user's upcoming mood autonomously based on their activities on OSNs without referring to extra devices or sensors. To achieve this, we first collect both online data and ground truth data about the mood of a set of Japanese users. We recruited participants for our experiment using an online survey company. Each user is requested to share their OSNs IDs and to report their *overall mood* at the end of the each day during the experiment by answering the following question: "How was your mood today, for the whole day in general?". We have an agreement from all participants on the condition to collect their data. Collecting the overall mood at the end of the day reflects our research objective of forecasting the overall mood of tomorrow and not the instantaneous mood.

We recruited users who are relatively active on Twitter, and Instagram or Foursquare. Facebook was omitted from our experiment because our data collection app was not approved in the Facebook review process. 383 users were chosen to participate in this study with 136 male and 247 female. All the 383 users have accounts on Twitter and Instagram. However, only 12 among them have a Foursquare account. In Japan, Foursquare is not among the highly used location-based applications. For this, we focused on Instagram as a source for location data. We ran the study for two months from May to July 2017. Each participant was requested to report daily about their mood on a Likert scale from 1 (very bad) to 5 (very good). Considering the amount of the collected data, we grouped the self-reported mood into 3 classes; positive (mood score: 4 and 5), negative (mood score: 1 and 2) and neutral (mood score: 3). A number of participants were deleted from the analysis using the below mentioned heuristics:

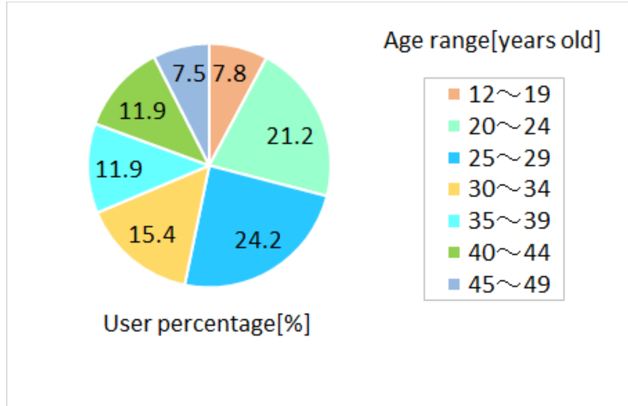


Fig. 3 The age range in the collected data

- Participants with private accounts or wrong OSNs IDs.
- Participants with one instance in each mood class.
- Participants with less than 10 days of online activity.
- Participants who entered the same mood every day during the experiment period (two months).

After data cleansing, the number of users in the dataset becomes 203.

4.2 Data Statistics

The majority of the participants were between 20 to 34 years old (see Fig. 3) with different professions.

Figure 4 shows the distribution of user profession; 16.4% Employee (office), 16% Housewives (husbands) and 14.3% were students.

4.3 Data Sample

We present in Fig. 5 and Fig. 6 a sample of SNS data. To respect copyright condition when we recruited the participants, we presented a sample of fictive data. For Instagram, users can edit and upload photos and short videos on the platform. Users can add a caption to each of their posts and use hashtags and location-based geotags to index these posts. However, for Twitter, the tweet is limited to 140 characters or less. A tweet may contain text including hyperlink

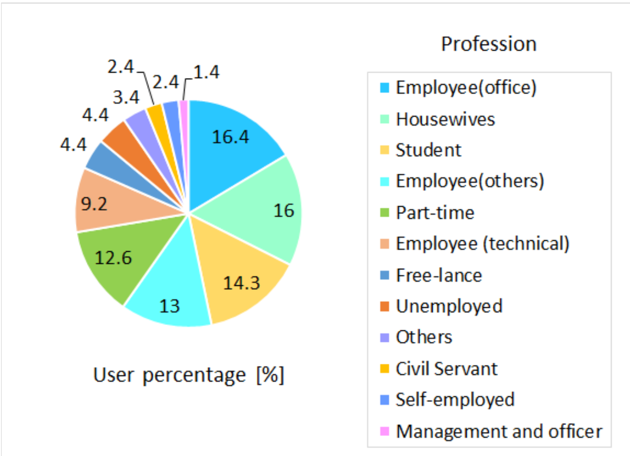


Fig. 4 Profession of the participants in our experiment

Fig. 5 Sample data: instagram

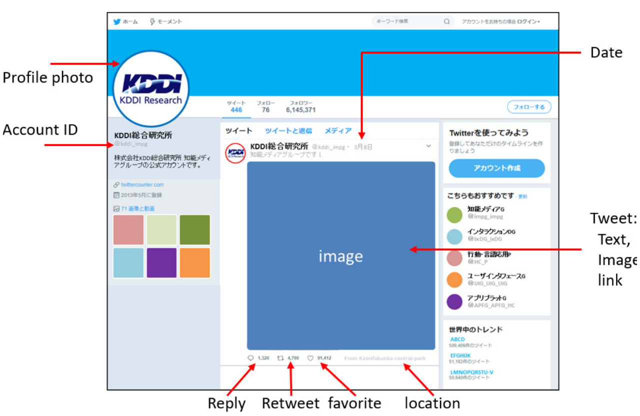


Fig. 6 Sample data: twitter

and picture. Tweets can be associated with a location, generating a tweet with geo-tag information.

5. Features Analysis

5.1 Correlation Analysis

The objective of this analysis is to check whether the extracted features are relevant for predicting the upcoming mood of user. We refer to Pearson correlation to test the statistical relationship between each feature and the user mood. Table 2 reports the results with statistically significant correlations adjusted using Bonferroni correction ($p < 0.05/N$ where N is the total number of features). Among 203 users, we found 127 users who have at least one feature with statistically significant correlations. For each feature, the mean, variance and the standard error of Pearson coefficient were calculated over the 127 users.

The results in Table 2 suggest the existence of statistically significant correlation between the mood and the selected features. Moreover, the result of Pearson coefficient r shows a high correlation between the selected set of features and the upcoming mood. The highest correlation was with geo-location feature with a Pearson value equals to 0.78. Instagram features extracted from images have, also, a high correlation with user's mood. This result is in alignment with our claim, that not only linguistic-based features that correlate with user's mood and that expanding the set of features to linguistic and non-linguistic features allows to have more correlated users than using only linguistic-based feature when forecasting their mood.

5.2 Features Stability

To test whether the set of feature per user change over time, we observe the users' feature sets selected from different time periods. For each user, we apply feature selection to select the highly correlated features from data collected in 40 days to form a feature set (F_1) and then we use the same method to select a feature set F_2 in 60 days long. We refer to Jaccard coefficient to measure the similarity of the two feature set F_1 and F_2. Jaccard coefficient is a value between 0 and 1. The higher the value is the more similar the two sets of features are. We show in Fig. 7 the distribution of the Jaccard coefficient. As shown in the figure, the half of users have a similarity coefficient larger than or equal to 0.7, which means their feature sets are unchanged during the period of experiment. This result implies that the extracted feature set changes slowly over time. We can say that our

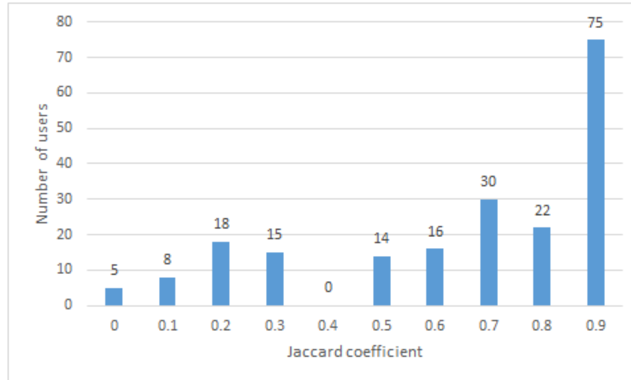


Fig. 7 Distribution of Jaccard coefficient

learning model can be applied for a relative long period of time without the need of frequent updates.

6. Prediction Evaluation

In this section, we focus on the correlated users to evaluate our proposal in terms of overall accuracy, precision and recall. We compare the performance with two linguistic-based methods.

1. Cyber-conventional: we extract the same features, described in [6], from a specific day and predict tomorrow's mood.
2. Linguistic-based method: we only use linguistic features to predict the upcoming user's mood. These features are related to text sentiment analysis. For this, we referred to Google Cloud Natural Language API to calculate negative/positive text count from Twitter posts and negative/positive caption count from Instagram posts. We use for this method, the first four features described in Table 2.

6.1 Personalized Mood Prediction

The goal, here, is to predict the upcoming mood of each user individually using their own collected data. The upcoming user mood of each day during the experiment is predicted and the average value is calculated over our 127 users. We use 4-fold cross validation to validate our model.

Table 3 summarizes the results in terms of accuracy, precision and recall. The proposed method shows an accuracy of 70% and a recall of 75%. This result outperforms the conventional schemes in all metrics. For example, the proposal performs better by 22 percentage points compared to the linguistic-based method in terms of recall. This result shows that user's mood can be forecast using cyber-social-physical features extracted from their online public data without using extra sensors or devices.

Moreover, we evaluate the performance of the proposal on a subspace of features to identify which combination is the best for prediction. The results are shown in Table 4. The results show that the combination of all features is the best for predicting tomorrow's mood. Also, the combination of social-physical features is the best in terms of precision and recall.

Table 3 Personalized: comparison between conventional methods and our proposal in forecasting user's mood

	Linguistic-based	Cyber-conventional	Proposal
Accuracy	0.55	0.62	0.70
Precision	0.59	0.67	0.73
Recall	0.53	0.69	0.75

Table 4 Personalized: proposal performance evaluation using a subspace of features

	All	Cyber-only	Social-only	Physical-only	Cyber-Social	Cyber-Physical	Social-Physical
Accuracy	0.70	0.55	0.52	0.54	0.60	0.62	0.61
Precision	0.73	0.53	0.56	0.58	0.63	0.67	0.67
Recall	0.75	0.57	0.61	0.60	0.64	0.60	0.69

6.1.1 The Effect of User Activeness on the Performance

The results about performance of the personalized system brings up a question about how much history data we need for each user to develop a decent predictive system for tomorrow's mood. To answer this question, we divided our users into two groups: active and non-active users.

- Active users: who have more than 30 days of online activity on OSN during the experiment. The number of active users in this group is 73.
- Non-Active users: who have less than 30 days of online activity on OSN during the experiment. The number of users in this group is 54.

We studied the performance of our personalized system in terms of accuracy, precision and recall averaged over the number of users in each group. The results are reported as follows.

(1) For Active users:

It is obvious that by considering only active users on OSNs, we can see a higher performance compared to the results shown in Table 3 of the personalized system averaged over all users. Also, the result in Table 5 shows that the proposal outperforms the conventional schemes; the recall value reaches 80% and a precision of 82%. This result shows that the activeness of the users will apply an affect on the prediction results. The more active the user is on OSNs, the higher accuracy we get in predicting the upcoming mood. However, people are usually active on one social network more than another. For this, we also show the result of Twitter-based features vs. Instagram-based features in Table 6. Instagram-based features outperform Twitter-based features, because these features expand over all spaces namely cyber, social and physical. Also, the accuracy decreased from 79% when extracting features from all SNS accounts to 74% when considering Instagram-based features.

(2) For Non-Active users:

Since users of this group have less than 30 instances dur-

Table 5 Active users: comparison between conventional methods and our proposal in forecasting user's mood

	Linguistic-based	Cyber-conventional	Proposal
Accuracy	0.58	0.69	0.79
Precision	0.64	0.71	0.82
Recall	0.59	0.70	0.80

Table 6 Active users: comparison between twitter-based and instagram-based features in forecasting user's mood

	Twitter-based	Instagram-based
Accuracy	0.70	0.74
Precision	0.72	0.76
Recall	0.70	0.75

Table 7 Non-active users: comparison between conventional methods and our proposal in forecasting user's mood

	Linguistic-based	Cyber-conventional	Proposal
Accuracy	0.45	0.49	0.56
Precision	0.51	0.53	0.59
Recall	0.50	0.50	0.57

Table 8 Non-active users: comparison between twitter-based and instagram-based features in forecasting user's mood

	Twitter-based	Instagram-based
Accuracy	0.49	0.54
Precision	0.52	0.57
Recall	0.49	0.55

Table 9 All users: comparison between conventional methods and our proposal in forecasting user's mood

	Linguistic-based	Cyber-conventional	Proposal
Accuracy	0.40	0.45	0.49
Precision	0.37	0.41	0.53
Recall	0.38	0.43	0.55

Table 10 All users: proposal performance evaluation using a sub-space of features

	All	Cyber-only	Social-only	Physical-only	Cyber-Social	Cyber-Physical	Social-Physical
Accuracy	0.49	0.40	0.34	0.38	0.42	0.46	0.46
Precision	0.53	0.44	0.31	0.47	0.48	0.52	0.51
Recall	0.55	0.47	0.39	0.45	0.49	0.53	0.52

ing the experiment, the data set for each mood class (positive, negative and neutral) is quite small that leads to a low performance (see Table 7) compared to the results in Table 3. However, compared to linguistic-based and cyber-conventional methods, our proposal is performing better in terms of accuracy, precision and recall.

As we did in the analysis of active users, we show the result of Twitter-based features vs. Instagram-based features in Table 8. For non-active users, the accuracy decreased by only 2 percentage points when using Instagram-only features. The effect of selecting features from one social network depends on the activeness of users on SNS. The more active the user is the more effect will be on the result. Our study shows that we needed to gather a certain amount of data about each user separately to get a decent personalized system.

6.2 Generalized Mood Prediction

The availability of individual data is not always guaranteed, especially at the initial phase of the system. To overcome data sparsity in the case of non-active users or when a new user joins the system, we opt for a generalized mood predictive system that uses the data collected from all users to predict the mood of a specific user. We evaluate the performance of the generalized system to see whether the generic system can be a solution to compensate the lack of data at the early stage of the system.

(1) All users:

We use all the collected data with 4-fold cross validation to build a one-for-all mood predictive system. The total number of users is 127. We report the result in terms of accuracy, precision and recall for tri-class classifier (positive, negative and neutral). Comparing the result of prediction, in Table 9 between conventional methods (linguistic-based and cyber-conventional) and our proposal, we can say that linguistic-based method performs slightly better than a random prediction; precision and recall value are 0.37 and 0.38, respectively. However, the proposed method can predict the upcoming mood with a precision of 0.53 and a recall of 0.55. Also, the performance of cyber-conventional method was lower by 12 percentage points than the proposal in terms of precision and recall. The result in Table 9 shows a weak performance for the generalized predictive system. This result comes from the characteristics of the mood; user-dependent. Moreover, we evaluate the performance of proposal when using different subspace of features to identify which combination is the best for prediction. The results in Table 10 show that when comparing the performance of each subspace alone, we found that cyber and physical subspaces are performing better than the social subspace. Same as the results found in personalized system, the combination of all features gives the best performance followed by the combination of cyber-physical subspace.

To summarize, we say that we couldn't get a good prediction from one-to-all model. For this, we tried to build a generalized system, but for a smaller group of users who have a specific characteristic in common in order to improve the accuracy of the generalized predictive system. There are some demographics that are correlated to mood such as age, gender and profession. In this work, we started with the most basic information i.e., gender. The age and profession are too granularly divided to generate a generic model from the data that we have. For this reason, we decided to build a generalized mood based on the gender information. The gender information divide the set of all-users into a smaller sets to make the prediction easier. We defined two study groups:

- Men: The number of men is our dataset is 50.
- Women: Then number of women is 77.

For each group, we trained a new generalized system per gender and report the result as follows:

(2) Men:

As shown in Table 11, considering a general classifier for men ameliorate the performance of the generalized system. The accuracy goes from 0.49 in one-for-all predictive system to 0.63 for men-only predictive system. When comparing the performance between conventional methods and our proposal, we can see a 10 and 5 percentage points of amelioration in accuracy, precision and recall over linguistic-based and cyber-conventional method, respectively.

Table 11 Men: comparison between conventional methods and our proposal in forecasting user's mood

	Linguistic-based	Cyber-conventional	Proposal
Accuracy	0.53	0.57	0.63
Precision	0.60	0.63	0.68
Recall	0.56	0.60	0.65

Table 12 Women: comparison between conventional methods and our proposal in forecasting user's mood

	Linguistic-based	Cyber-conventional	Proposal
Accuracy	0.60	0.64	0.68
Precision	0.58	0.65	0.74
Recall	0.48	0.58	0.70

(3) Women:

The amelioration in terms of accuracy is almost 20 percentage points compared to the one-for-all system. Also, our proposal outperforms the conventional methods in all metrics. For recall results, the proposal achieves 0.70 compared to 0.48 for linguistic-based method and 0.58 for cyber-conventional method. We can say that women predictive system (accuracy 68%) performs better than men predictive system (accuracy 63%). This is justified by the fact that women are the more emotionally expressive gender [20]. The results shown in Table 11 and Table 12 are in align with our claim that considering a general classifier for a set of users ameliorates the performance of the generalized system. Of course, the performance of the generalized systems (i.e., men, women) could not reach the performance of the personalized system but, at least we can refer to such system when the user has no enough history data to build his own personalized system. To conclude, using the combined data from a group of users who have a specific characteristic (i.e., gender) in common could be a solution to overcome data sparsity at the initial phase of the system. However, to achieve a decent prediction results a personalized system is essential when having enough history data.

To support this conclusion, we conducted an experiment considering users who have more than 30 days of online activities on OSNs (73 users). We compared the average performance (F1-measure) over all users between personalized and generalized (per gender) systems. We plotted the result in Fig. 8. As we can see from the personalized system in red line, the more data we have the better the performance is. At early stage, the generalized system is performing better than the personalized system, because of lack of information about each user. However, when time elapses, the personalized system outperforms the generalized system after collecting enough data about the user. From Fig. 8, we can say that for the users in this simulation, the turning point is located at day 28 when the personalized system outperforms the generalized system. The position of this turning point over time differs from one user to another. Identifying the turning point defines when the system should switch from the generalized (gender-based) system to the personalized one when forecasting a user's mood. One method to do this

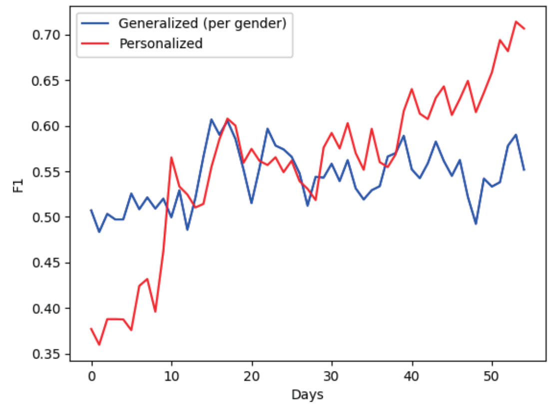


Fig. 8 Comparison between gender-based and personalized systems in terms of F1-measure

is to wait until the performance of the personalized system outperforms the generalized system for a specific time window before switching to the personalized prediction. The study of the optimal time is beyond the scope of this paper.

7. Conclusion and Future Work

In this paper, we implement a personalized and generalized system for forecasting the mood based on online data on user's OSN accounts. We have shown that predicting tomorrow's mood can be done using his cyber, social and physical features extracted, today, from the user's public information shared on online social networks (OSNs). This task is different from predicting the current user's mood. To do so, we conducted an experiment for two months to collect ground truth data about user's mood and their OSNs public data.

The result of a personalized predictive system showed a high performance of the proposed method in terms of accuracy (70%), precision (73%) and recall (75%). Our results suggested that the performance of the personalized system depends on the activeness of the user on OSNs. For users with 30 days of history data, we achieved 82% in terms of precision. We also showed that a general predictive system for a set of users (i.e; women, men) can be used at the early stage of the system when no enough data is available to build a personalized system for each user separately.

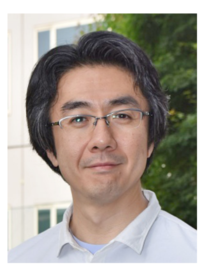
As mentioned in Sect. 5.1, not all the users in our experiment showed a statistically significant correlation with the extracted features. Expanding the analysis to multiple data sources can be a solution to this limitation. That is why, as future work, we want to expand our experiments by analyzing data of each user gathered from OSNs and smartphone sensors. Moreover, in this work, the mood was treated as a tertiary state (e.g., negative, positive and neutral). However, there are multiple ways to extend the proposed method. We hope future work will develop these extensions and achieve a precise forecast for a multi-class state of user's mood. We, also, plan to study the effect of history data on forecasting user's mood and to expand our set of features by considering more social networks.

References

[1] Wikipedia, "Mood (psychology) — Wikipedia, the free encyclopedia." https://en.wikipedia.org/wiki/Mood_(psychology), note = "[Online; accessed 06-March-2019]," 2019.
[2] G. Adomavicius and A. Tuzhilin, "Context-aware recommender systems," Proceedings of the 2008 ACM Conference on Recommender Systems, RecSys '08, New York, NY, USA, pp.335–336, ACM, 2008.
[3] M. Roshanaei, R. Han, and S. Mishra, "Having Fun?: Personalized Activity-Based Mood Prediction in Social Media," pp.1–18, Springer International Publishing, Cham, 2017.
[4] S. Shah, J.N. Teja, and S. Bhattacharya, "Towards affective touch interaction: predicting mobile user emotion from finger strokes," Journal of Interaction Science, vol.3, no.1, p.6, Nov. 2015.
[5] M. Roshanaei, R. Han, and S. Mishra, "Features for mood prediction in social media," 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp.1580–1581, Aug. 2015.
[6] J.A. Lee, C. Efstratiou, and L. Bai, "Osn mood tracking: Exploring the use of online social network activity as an indicator of mood changes," Workshop on Mental Health Sensing and Intervention in conjunction with UBICOMP'16, pp.1171–1179, September 2016.
[7] S.A. Taylor, N. Jaques, E. Nosakhare, A. Sano, and R. Picard, "Personalized multitask learning for predicting tomorrow's mood, stress, and health," IEEE Transactions on Affective Computing, vol.PP, no.99, pp.1–1, 2017.
[8] A. Mikus, M. Hoogendoorn, A. Rocha, J. Gama, J. Ruwaard, and H. Riper, "Predicting short term mood developments among depressed patients using adherence and ecological momentary assessment data," Internet Interventions, vol.12, pp.105–110, 2018.
[9] M. Mehraei and N.I. Akcay, "Pleasure, arousal, and dominance mood traits prediction using time series methods," IAFOR Journal of Psychology and the Behavioral Sciences, vol.3, no.1, Aug. 2017.
[10] V. Hollis, A. Konrad, A. Springer, M. Antoun, C. Antoun, R. Martin, and S. Whittaker, "What does all this data mean for my future mood? actionable analytics and targeted reflection for emotional well-being," Human-Computer Interaction, vol.32, no.5-6, pp.208–267, 2017.
[11] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft, "Our twitter profiles, our selves: Predicting personality with twitter," Proceedings of the Third International Conference on Social Computing (SocialCom) and the Third International Conference on Privacy, Security, Risk and Trust (PASSAT), pp.180–185, IEEE, Oct. 2011.
[12] B. Ferwerda, M. Schedl, and M. Tkalcic, "Predicting personality traits with instagram pictures," Proceedings of the 3rd Workshop on Emotions and Personality in Personalized Systems 2015, EMPIRE '15, New York, NY, USA, pp.7–10, ACM, 2015.
[13] M.D. Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," ICWSM, ed. E. Kiciman, N.B. Ellison, B. Hogan, P. Resnick, and I. Soboroff, The AAAI Press, 2013.
[14] M. De Choudhury, S. Counts, and E. Horvitz, "Predicting postpartum changes in emotion and behavior via social media," Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, New York, NY, USA, pp.3267–3276, ACM, 2013.
[15] L.T. Nguyen, P. Wu, W. Chan, W. Peng, and Y. Zhang, "Predicting collective sentiment dynamics from time-series social media," Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, WISDOM '12, New York, NY, USA, pp.6:1–6:8, ACM, 2012.
[16] R. Likamwa, Y. Liu, N.D. Lane, and L. Zhong, "Moodscope: Building a mood sensor from smartphone usage patterns," MobiSys 2013 - Proceedings of the 11th Annual International Conference on Mobile Systems, Applications, and Services, pp.465–466, 2013.
[17] Y. Suhara, Y. Xu, and A.S. Pentland, "Deepmood: Forecasting depressed mood based on self-reported histories via recurrent neural networks," Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017, pp.715–724, 2017.
[18] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of twitter data," Proceedings of the Workshop on Languages in Social Media, LSM '11, Stroudsburg, PA, USA, pp.30–38, Association for Computational Linguistics, 2011.
[19] C. Dhahri, K. Matsumoto, and K. Hoashi, "Cyber-social-physical features for mood prediction over online social networks," 2017 Forum on data engineering and information management (DEIM), March 2017.
[20] A.M. Kring and A.H. Gordon, "Sex differences in emotion: expression, experience, and physiology," Journal of personality and social psychology, vol.74, no.3, pp.686–703, 1998.

Dhahri Chaima received the B.E. degree in Computer Science from INSAT University (Tunisia) in 2010 and M.E. and Ph.D. degree from Keio University (Japan) in 2013 and 2016. During her studies, her research was about the application of machine learning in wireless communication. She has been working at KDDI Research Inc.since 2016 as a Researcher at the Intelligent Media Laboratory. Her work is mainly about the analysis of SNS data and phone sensor data for mood prediction and recommendation systems.

Kazunori Matsumoto received the B.E. and M.E. degrees in information science from Kyoto University, Kyoto, Japan, in 1984 and 1986, and Ph.D. degree from Ritsumeikan University, Kyoto, in 2009 respectively. He has been with KDDI Research, Inc. Saitama, Japan, since 1986, and is currently a Research Manager of Intelligent Media Processing Group. His current research interests include multimedia content analysis and human computer interaction.



Keiichiro Hoashi received the B.E. and M.E. degree in Information Science from Waseda University in 1995 and 1997, and Ph.D. degree from Waseda University in 2007. He has been working at KDDI Research Inc. since 1997, and is currently the Senior Manager of the Intelligent Media Laboratory. He has been mainly engaged in the research areas of multimedia information retrieval, machine learning, and natural language processing. He was awarded the FIT 2004 Young Researcher award, and also worked as a part-time lecturer at Waseda University from 2001 to 2005. He was also selected as a member of the Sido 2015 Next Innovator program sponsored by METI (Ministry of Economy, Trade and Industry) of Japan. Currently members of ACM, and IPSJ.