Towards Unobtrusive Emotion Recognition for Affective Social Communication

Hosub Lee, Young Sang Choi, Sunjae Lee, and I. P. Park
Intelligent Computing Laboratory, Future IT Research Center
Samsung Advanced Institute of Technology (SAIT), Samsung Electronics Co., Ltd.
Nongseo-dong, Giheung-gu, Yongin-si, Gyeonggi-do, Korea 446-712
{horus.lee, macho, sunjae79.lee, ilpyung.park}@samsung.com

***Abstract*—Awareness of the emotion of those who communicate with others is a fundamental challenge in building affective intelligent systems. Emotion is a complex state of the mind influenced by external events, physiological changes, or relationships with others. Because emotions can represent a user's internal context or intention, researchers suggested various methods to measure the user's emotions from analysis of physiological signals, facial expressions, or voice. However, existing methods have practical limitations to be used with consumer devices, such as smartphones; they may cause inconvenience to users and require special equipment such as a skin conductance sensor. Our approach is to recognize emotions of the user by inconspicuously collecting and analyzing user-generated data from different types of sensors on the smartphone. To achieve this, we adopted a machine learning approach to gather, analyze and classify device usage patterns, and developed a social network service client for Android smartphones which unobtrusively find various behavioral patterns and the current context of users. Also, we conducted a pilot study to gather real-world data which imply various behaviors and situations of a participant in her/his everyday life. From these data, we extracted 10 features and applied them to build a Bayesian Network classifier for emotion recognition. Experimental results show that our system can classify user emotions into 7 classes such as happiness, surprise, anger, disgust, sadness, fear, and neutral with a surprisingly high accuracy. The proposed system applied to a smartphone demonstrated the feasibility of an unobtrusive emotion recognition approach and a user scenario for emotion-oriented social communication between users.**

***Index Terms*—Affective computing, Computer mediated communication, Emotion recognition, Machine intelligence, Supervised learning**

I. Introduction

SOCIAL networking service (SNS) like Twitter, Facebook or Google+ is an online service platform which aims for building and managing social relationships between people. On the SNS, users represent themselves in various ways such as profiles, pictures, or text messages, and interact with other people including their acquaintances. By using various SNS applications on mobile devices, furthermore, users can share ideas, interests or activities at any time, and anywhere with their friends.

More specifically, a SNS user often expresses her/his feeling or emotional state directly with emoticons or indirectly with written text, thereby their followers respond to her/his message more actively or even empathize with them; In psychology, this kind of phenomenon is called the emotional contagion [1]. That is, communication between users would be newly induced or enriched by the sharing of emotion. We defined these kind of communicational activities as affective social communication as depicted in Figure 1.

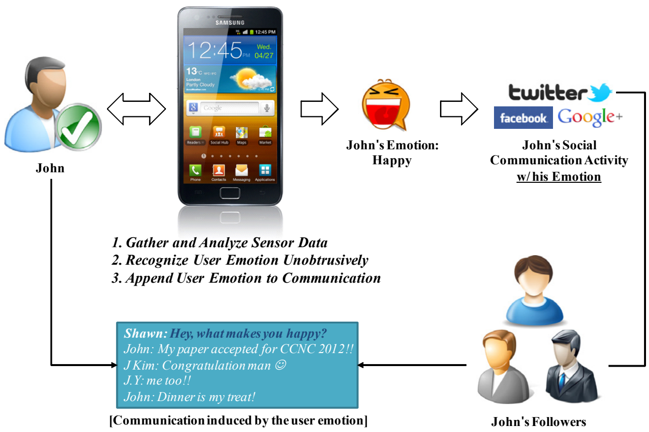


Fig. 1. Conceptual Diagram of Affective Social Communication.

However, some SNS users hardly show what they feel; therefore their friends cannot sense or react to their emotional states appropriately. This is probably because they are unfamiliar with the expression of emotion or are not aware of their own emotions enough. Possible solution for this problem is adopting emotion recognition technologies which are being extensively studied by the affective computing research society to determine emotions of the user. Existing emotion recognition technologies can be divided into three major categories depending on what kinds of data is analyzed for recognizing human emotion: physiological signals, facial expressions, or voice. Physiological emotion recognition shows acceptable performance but has some critical weaknesses that prevent its widespread use; they are obtrusive to users and need special equipment or devices. For example, if we want to recognize user emotion from their bio-signal, the user should wear specialized equipment such as a skin conductance sensor, blood pressure monitor, or electrocardiography (ECG)-derived heart rate monitor on their body [2]. These devices are obviously not only intrusive to users, but involve additional expense. On the other hand, emotion recognition using facial expressions or speech has limitations on its usage, because the device needs to be positioned in front of the face of the user or should always listen to the voice of the user, which does not happen frequently in daily use such as text messaging.

In this paper, we present a machine learning approach to recognize emotional states of a smartphone user without any inconvenience or extra cost for equipping additional sensors. To perceive current emotion of the user, we gathered various types of sensor data from the smartphone. These sensor data can be categorized into two types such as behavior and context of the user, and it was collected while the user uses a certain application on her/his smartphone. As the proof of concept, we developed an Android application named affective twitter client (AT Client) which collects abovementioned sensor data whenever the user sends a text message (i.e., tweet) to the Twitter. Via the pilot study for two weeks, we collected 314 dataset including self-reported emotional states of the user, and utilized it to build Bayesian Network classifier which is a powerful probabilistic classification algorithm based on the Bayes theorem [3]. Through repetitive experiments including preprocess, feature selection, and inference (i.e., supervised learning), our classifier can classify each tweet written by the user into 7 types with a satisfactory accuracy of 67.52% on average. Classifiable types of emotions include Ekman's six basic emotions [4] and one neutral state.

We believe that the presented method can unobtrusively recognize emotions of individual smartphone users, and it will affectively enrich social communication by sharing recognized emotion among users. Also, we believe that affective intelligence in a broader sense can provide differentiating values to smartphone manufacturers who face with fierce competition for the market.

II. Background

A. Affective Computing

Affective computing can be defined as the study of technologies to recognize, model, and express human affective phenomena such as emotion, moods, attitudes, and personality traits. The ultimate goal of this research area is to enable machines to understand emotional states of humans and give an appropriate reaction for those emotions, thereby enhancing human's emotion and feelings. R. W. Picard first proposed a fundamental concept of affective computing and its technical descriptions such as signal-based representation of emotions, human emotion recognition, and methods to build a emotion model to synthesize emotions in computers [5].

B. Emotion Recognition

Many affective computing and human computer interaction (HCI) researchers have suggested various methods to sense and recognize human emotions.

C. Peter *et al.* proposed wearable system architecture for collecting emotion-related physiological signals such as heart rate, skin conductivity, and a skin temperature of users [6]. They also developed a prototype system, consisting of a glove with a sensor unit, and a base unit for receiving the data transmitted from the sensor unit. Using a robust data evaluation method, they validated that collected data are available and reliable. Although their system is easy to use and suitable for mobile computing, it has a limitation that users should wear a glove.

S. V. Ioannou *et al.* demonstrated an emotion recognition system through evaluation of facial expressions [7]. To recognize emotions, they developed a neuro-fuzzy network based on rules which have been defined via analysis of facial animation parameters (FAPs) variations of users. With experimental real data, they also showed acceptable recognition accuracy of higher than 70%. A. Batliner *et al.* presented an overview of the state of the art in automatic recognition of emotional states using acoustic and linguistic parameters [8]. They summarized core technologies such as corpus engineering, feature extraction, and classification have been used for building emotion recognition systems via the speech analysis. Even though these approaches are novel and promising, users need to have an expressive face and utter sounds or words respectively.

III. Unobtrusive Emotion Recognition

A. Related Work

Unlike existing emotion recognition technologies outlined above, we are focusing on the situation that users use a smartphone with an expressionless face or silence, and they do not want to feel any burden associated with the recognition process. There have been attempts to satisfy these requirements; recognize emotions of the user inconspicuously with a minimum cost.

M. Poh *et al.* developed a method for measuring multiple physiological signals using a general webcam [9]. They applied independent component analysis (ICA) on the color channels in facial image caught on a webcam, and extracted various vital signs such as a heart rate (HR), respiratory rate, and HR variability. To prove lightness and applicability of their approach, they utilized a commonly used webcam as a sensing unit. Their approach showed significant potentials for affective computing, because there is a close correlation between the bio-signal such as a HR and emotion, and it does not require any attention of users. C. Epp *et al.* proposed a solution to determine emotions of computer users by analyzing the rhythm of their typing patterns on a standard keyboard [10]. By conducting a field study, they gathered various keystrokes of participants and build a C4.5 decision tree classifier for 15 emotional states. Through experiment, they successfully modeled six emotional states including confidence, hesitance, nervousness, relaxation, sadness, and tiredness with classification accuracies ranging from 77.4% to 87.8%. As a similar methodology of ours, this work showed a promising result for unobtrusive emotion recognition.

B. Methodology

To determine the current emotion of users, we adopted a machine learning approach consisting of the data collection, data analysis, and classification process. For these tasks, we used a popular machine learning software toolkit named Weka [11].

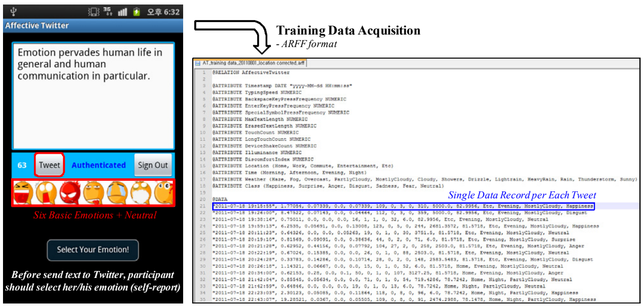


Fig. 2. Data Collection Process. A participant sends a tweet to Twitter when she/he feels a certain emotion, and AT Client collects associated sensor data with the self-reported user emotion. These data are utilized for the data analysis process.

Data collection process is visualized as a Figure 2. In this process, we gathered various sensor data from the smartphone when a participant use AT Client installed on their device. In more detail, if a participant feels a specific emotion at the certain moment in their everyday life, they would write some short text messages and were asked to report their current emotion via AT Client. Meanwhile, AT Client collected various sensor data during this period. It is also possible that some users write a tweet without any emotions; this emotional state might be a neutral, and we also took this as training data

TABLE I
LIST OF FEATURES

Attribute Name	Description
Typing Speed	All of these features are generated according to typing behavior on the default Android widget named EditText. From these data, we can infer habits of users in writing text messages. All values are *numerical*.
Backspace Key Press Freq.	
Enter Key Press Freq.	
Special Symbol Press Freq.	
Maximum Text Length	
Erased Text Length	
Touch Count	It means how many times user invokes embedded edit window in EditText to perform various functions like cursor movement, word selection, and copy & paste. All values are *numerical*.
Long Touch Count	
Device Shake Count	It means how much the device is shaken; it has a *numerical* value.
Illuminance	It means ambient brightness; it has a *numerical* value.
Discomfort Index (DI)	It is calculated based on the formula of Thom; $DI = 0.4 \times (Ta + Tw) + 15$ where Ta = dry-bulb temperature (F), Tw = wet-bulb temperature (F); it has a *numerical* value.
Location	Home, Work, Commute, Entertain., Etc
Time	Morning, Afternoon, Evening, Night
Weather	14 weather conditions defined by the Google weather

Class Attribute (7 emotional states): Happiness, Surprise, Anger, Disgust, Sadness, Fear, Neutral

because the neutral emotion itself is a frequently observed state in emotion recognition [12]. From gathered sensor data, we extracted 14 features which seem to have potential correlations with the emotion. All of these features are formalized as an attribute-relation file format (ARFF) for Weka. List and descriptions of features are summarized in Table I.

Next, we analyzed collected training data using Weka. As the first step, we discretized features which have a continuous numerical value such as a Typing Speed. This procedure is necessary for making all features suitable for numerical and statistical computation on the Weka (i.e., preprocess step). Then, we ranked all features by measuring information gain with respect to each class. Information gain is one of the most popular metrics of association and correlation among several random variables [13]. Based on the result of attribute evaluation, we finally selected 10 features which have more strong correlation with the emotion to build an inference model (i.e., feature selection step). Shadowed cells of Table I are selected features and reasonable explanations for this feature selection are presented in the experimental study section. By the repetitive experiments using 10-fold cross validation, we finally chose Bayesian Network classifier as an inference model for our system, because it showed highest classification accuracy among other machine learning classifiers such as a Naïve Bayes, decision tree, or neural network. Sample Bayesian Network is drawn at the Figure 3.

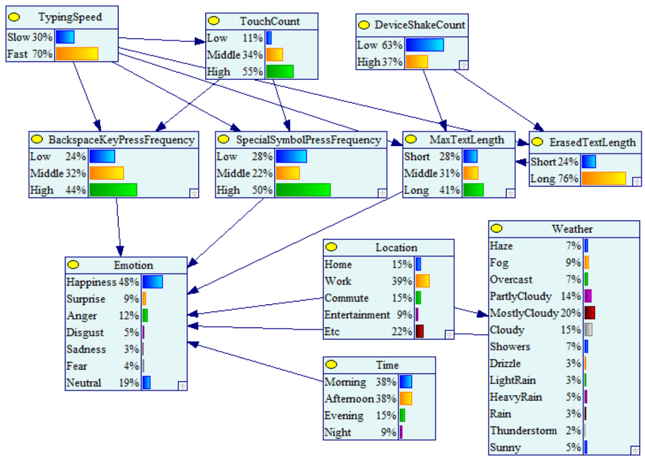


Fig. 3. Sample Bayes. Network drawn by *GeNIe* (http://genie.sis.pitt.edu/) software

At last, our inference model was continuously updated by the unknown real-world data and classified it into 7 emotional states (i.e., learning & inference step).

C. System Architecture

A block diagram of AT Client is shown in Figure 4. Data Aggregator gathers various data from internal/external sensors: it collects sensor data or information like coordinates of touch positions, degree of movement of the device, current location, or weather received from web through running software in smartphones. Then, Data Aggregator divides data into two classes such as user behavior-related and context-related data, and delivers these data to User Behavior Analyzer and User Context Analyzer respectively.

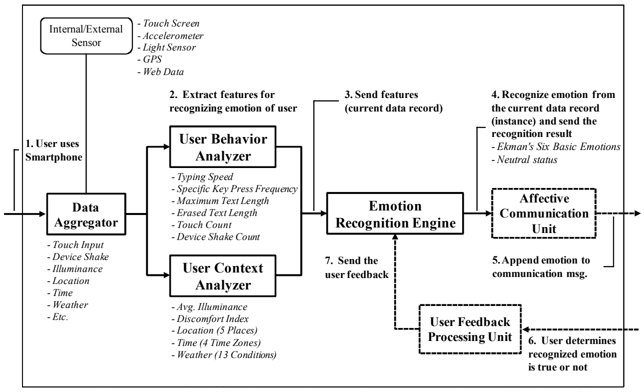


Fig. 4. Architecture of Affective Twitter Client. Parts shown in dotted line are being developed (*Work in progress*).

By analyzing various user input to the edit area of AT Client, User Behavior Analyzer extracts various features related to typing behavior of the user: typing speed, frequency of pressing a specific key, maximum text length, erased text length, and touch count. Additionally, it counts how many times the device is shaken when the user types a tweet via AT Client. At the same time, User Context Analyzer calculates features describing the current environmental conditions around the user: average brightness, discomfort index, location, time zone, and weather condition. Then, both User Behavior Analyzer and User Context Analyzer send all of these features (i.e., one data record in training dataset) to Emotion Recognition Engine.

Emotion Recognition Engine including Bayesian Network classifier categorizes incoming data into 7 types of emotions: happiness, surprise, anger, disgust, sadness, fear, and neutral, and delivers the result of emotion recognition to Affective Communication Unit. Affective Communication Unit utilizes recognized user emotion for enhancing affective experience of Twitter users; for example, it marks the emotional state of the user to her/his tweet or personal profile page. Additionally, User Feedback Processing Unit examines whether the user decides emotion recognition is accurate or not. To improve classification accuracy, User Feedback Processing Unit sends back this result to Emotion Recognition Engine. That is, Emotion Recognition Engine can continuously re-learn with the training dataset verified by the user.

D. Implementation

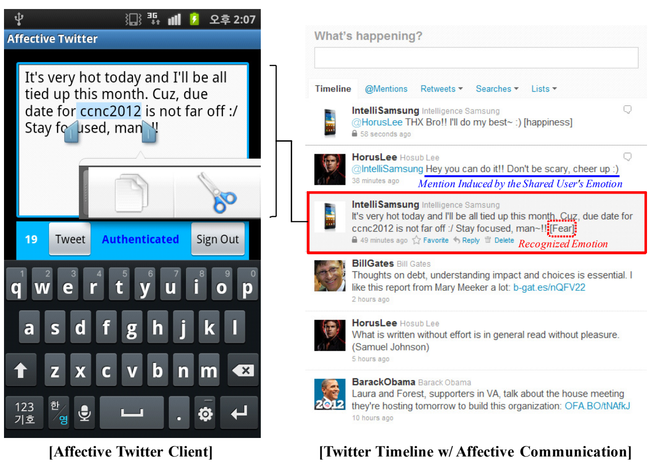


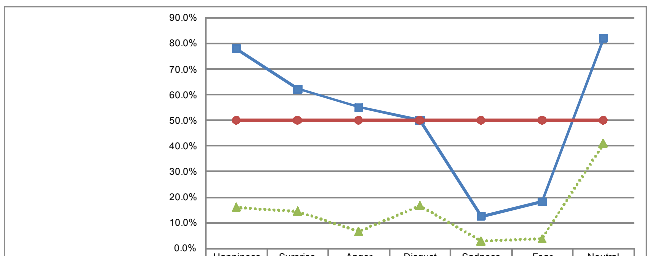
Fig. 5. Affective Social Communication.

As data collection software, we developed a Twitter client application named AT Client for a state of the art Android smartphone, Samsung Galaxy S II which has various built-in sensors. We implemented AT Client for the Android 2.3.3 platform (i.e., Gingerbread), and used Google APIs to get a location information of the user and Twitter4J [14] to access Twitter service respectively. To prevent any inconvenience or unfamiliarity to the participant, furthermore, we designed and implemented user interface of this prototype application to be simple and intuitive just like an official Twitter application. Figure 5 shows a screenshot of AT Client and a user scenario of affective social communication.

IV. Experimental Study

Through a pilot study for two weeks, we gathered 314 real-world dataset from a participant and utilized it to validate our emotion recognition approach by measuring classification accuracy. One subject from our research group members (one male in his 30s) was recruited for the pilot study; he wrote tweets and indicated his emotional state whenever he feels a certain emotion in his daily life.

Before classification, we evaluated all features via information gain attribute evaluation algorithm, and excluded four features in the lower ranks. Excluded features are frequency of enter key pressing, number of long touches, ambient light, and the discomfort index. In regard to enter key pressing and long touch, a participant hardly had begun a new line (i.e., enter key pressing) and done copy & paste (i.e., long touch) in editing a tweet message via AT Client; on the other hand, both brightness and discomfort index had little effect on his emotional state because he mostly stayed at the consistently bright and air-conditioned indoor places. For these reasons, we selected remaining 10 features as primary attributes of training data for our Bayesian Network classifier. The feature which has the highest correlation to emotions was the speed of typing; in addition, lengths of inputted text, shaking of the device, or user location were also important features for emotion recognition.



	Happiness	Surprise	Anger	Disgust	Sadness	Fear	Neutral
Classification Accuracy (AT Client)	78.0%	62.2%	55.0%	50.0%	12.5%	18.2%	82.0%
Classification Accuracy (Random)	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%
Ratio of Observed Emotions	15.9%	14.3%	6.4%	16.6%	2.5%	3.5%	40.8%

Fig. 6. Experimental Results for Unobtrusive Emotion Recognition.

Figure 6 shows experimental results for our approach. It showed average classification accuracy of 67.52% for 7 emotions. For emotions such as happiness, surprise, and neutral, our approach performed appreciably better than chance (i.e., 50%), and in the case of anger and disgust, it also outperformed at least random classification. Classification accuracy is irregular for emotion types, but we found a general correlation between the number of observation cases (green line) and classification accuracy (blue line). This finding means that low accuracy in classification for sadness and fear may be improved with additional data collection.

Therefore, we plan to gather more dataset from negatively excited participants. From these additionally collected data, we may need to discover new features for recognizing sadness and fear more accurately. Detailed classification results are tabularized as the Table II. It shows information about actual and predicted classification done by the proposed system; in other words, it indicates which items are misclassified into which classes.

TABLE II
CONFUSION MATRIX OF EXPERIMENTAL RESULTS

H	Su	A	D	Sa	F	N	Classified as
39	2	0	8	0	0	1	H
1	**28**	0	2	0	3	11	Su
3	1	**11**	5	0	0	0	A
12	3	2	**26**	0	0	6	D
1	0	0	3	**1**	0	3	Sa
0	2	0	1	0	**2**	6	F
0	13	0	8	0	2	**105**	N

H: Happiness, Su: Surprise, A: Anger, D: Disgust, Sa: Sadness, F: Fear, N: Neutral

V. CONCLUSION & FUTURE WORK

In this paper, we proposed an unobtrusive emotion recognition approach for affective social communication on mobile devices. With the development of an Android application named affective twitter client (AT Client), we gathered various real-world data and information from the Android smartphone equipped with diverse built-in sensors. By analyzing these dataset, we discovered 10 features related to the emotional state of the human user; these features are mainly divided into user behavioral patterns (e.g., typing speed) and the user context (e.g., location). With the training dataset including selected 10 features, we built Bayesian Network classifier and it showed classification accuracy of 67.52% on average for 7 emotional states: happiness, surprise, anger, disgust, sadness, fear, and neutral.

As the future work, we plan to improve the accuracy of classification for such negative emotions as sadness and fear by gathering and analyzing more training data from at least 5 more participants. For the improvement of overall classification performance, we will investigate new features associated with user behavior. For example, speed or intensity of a continual touch (e.g., drag operation) on AT Client would be a meaningful feature, because some users move their finger more violently on the touch screen when they are emotionally excited. To prove the usefulness of our approach, we will also conduct a user study including the discovery of the most user-friendly way in showing recognized user emotion via AT Client or Web. We believe that the proposed technology can make mobile device users affectively positive through automatic recognition and sharing of their emotion.

ACKNOWLEDGMENT

The authors would like to thank a participant who provided the training data we used. We used emoticon images from http://www.2s-space.com to develop the application.

REFERENCES

[1] J. Guillory, J. Spiegel, M. Drislane, B. Weiss, W. Donner, and J. T. Hancock, "Upset Now?: Emotion contagion in distributed groups," in *Proc. ACM CHI Conf. on Human Factors in Computing Systems,* Vancouver, 2011, pp. 745–748
[2] R. B. Knapp, J. Kim, and E. André, "Physiological signals and their use in augmenting emotion recognition for human–machine interaction," in *Emotion-Oriented Systems*, Pt. 2, Springer-Verlag, 2011, pp. 133–159
[3] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," in *Machine Learning*, vol. 29, Kluwer Academic Publishers, 1997, pp. 131–163
[4] P. Ekman, "Universals and cultural differences in facial expressions of emotions," in *Nebraska Symposium on Motivation*, vol. 19. University of Nebraska Press, 1972, pp. 207–283
[5] R. W. Picard, *Affective Computing*, Cambridge, MIT Press, 2000
[6] C. Peter, E. Ebert, and H. Beikirch, "A wearable multi-sensor system for mobile acquisition of emotion-related physiological data," in *Affective Computing and Intelligent Interaction*, vol. 3784, Springer-Verlag, 2005, pp. 691–698
[7] S. V. Ioannou, A. T. Raouzaiou, V. A. Tzouvaras, T. P. Mailis, K. C. Karpouzis, and S. D. Kollias, "Emotion recognition through facial expression analysis based on a neurofuzzy network," in *Neural Networks*, vol. 18 (4), Elsevier, 2005, pp. 423–435
[8] A. Batliner, B. Schuller, D. Seppi, S. Steidl, L. Devillers, L. Vidrascu, T. Vogt, V. Aharonson, and N. Amir, "The automatic recognition of emotions in speech," in *Emotion-Oriented Systems*, Pt. 2, Springer-Verlag, 2011, pp. 71–99
[9] M. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomedical Engineering*, vol. 58 (1), Jan. 2011, pp. 7–11
[10] C. Epp, M. Lippold, and R. L. Mandryk, "Identifying emotional states using keystroke dynamics," in *Proc. ACM CHI Conf. on Human Factors in Computing Systems,* Vancouver, 2011, pp. 715–724
[11] Weka 3, http://www.cs.waikato.ac.nz/ml/weka/
[12] R. Tato, R. Santos, R. Kompe, and J. M. Pardo, "Emotional space improves emotion recognition," in *Proc. Int'l Conf. on Spoken Language Processing*, Denver, 2002, pp. 2029–2032
[13] M. Martínez-Morales, N. Cruz-Ramírez, J. L. Jiménez-Andrade, and R. Garza-Domínguez, "Bayes-N: An algorithm for learning bayesian networks from data using local measures of information gain applied to classification problems," in *3rd Mexican Int'l Conf. on Artificial Intelligence*, Mexico City, 2004, pp. 527–535
[14] Twitter4J, Unofficial java library for the Twitter API, http://twitter4j.org/