Emotional arousal estimation while reading comics based on physiological signal analysis

Mizuki Matsubara
Osaka Prefecture University
1-1, Gakuen-cho, Naka,
Sakai, Osaka, Japan
matsubara@m.cs.osakafu-u.ac.jp

Olivier Augereau
Osaka Prefecture University
1-1, Gakuen-cho, Naka,
Sakai, Osaka, Japan
augereau.o@gmail.com

Charles Lima Sanches
Osaka Prefecture University
1-1, Gakuen-cho, Naka,
Sakai, Osaka, Japan
charles-ls@hotmail.fr

Koichi Kise
Osaka Prefecture University
1-1, Gakuen-cho, Naka,
Sakai, Osaka, Japan
kise@cs.osakafu-u.ac.jp

ABSTRACT

Estimating emotions while reading enables new services such as comic recommendation. Most of existing emotion estimation systems employ bulky devices. Furthermore, few applications have been developed for analyzing the emotions while reading. The purpose of our research is to develop a method for estimating emotions while reading. As the target of reading, we select comics, which stimulate emotions often more than other types of documents. As we want our system to be easily usable, we selected sensors embedded in a wristband and an eye tracker. Emotions can be described by two dimensions called emotional valence and arousal. As a first step, we propose in this paper to estimate the emotional arousal. We analyze the electrodermal activity, blood volume pulse, heart rate, skin temperature and pupil diameter of a subject to estimate if the reader feels a high or low arousal while reading. Our experiment shows that for some participants, the arousal can be estimated accurately.

CCS Concepts

•Human-centered computing → Ubiquitous and mobile computing;

Keywords

comic, emotion, reading, physiological signal

MANPU '16, December 04 2016, Cancun, Mexico

DOI: http://dx.doi.org/10.1145/3011549.3011556

1. INTRODUCTION

Comics have become more and more popular all over the world. E-book technology enables readers to carry many comics in their devices such as mobile phones, and to browse them with on-line services. Since it is difficult to find an enjoyable comic from such a large quantity of comics, a search engine and a recommendation system are required.

Generally, users search comics with the title, the name of the author and the publishing company such as the on-line search engine on the web page "Comic plus" [8] provided by Kodansha. Using such search engines is problematic if users do not know words to input. If users want to read invigorating comics when they are down in the dumps, it is not possible with the existing engines to search such comics.

In order to search the comics eliciting wanted emotions, the comics must be tagged with emotions. Tagging comics with emotions also enables a system to recommend comics according to the readers' emotions. A difficult problem of emotion based tagging is the readers' emotions are not always the same even if the content is the same. As an example, a comic which makes a reader sad might make another reader displeasure. Tagging comics with emotions enables to find the similarities of the elicited emotions between the readers. A reader's emotion while reading a comic can be estimated by analyzing the tags of similar readers. For example, a reader's emotion is estimated to be happy if most of the tags of the similar readers are happy. A simple way to obtain such tags is to ask the readers to tag manually their emotions after reading. However, it is laborious. Thus, the emotions must be tagged automatically by emotion estimation.

The emotions are elicited by various types of stimuli from sensory nerves and affect readers' bodies. Much work for estimating the emotions has been done by measuring the stimulated sensory nerves and the affected body. However, most of the existing methods employ bulky devices such as electroencephalogram [2]. Furthermore, most of the existing work has intended to analyze behavior of watching videos [5] and little work has been done to analyze behavior of reading

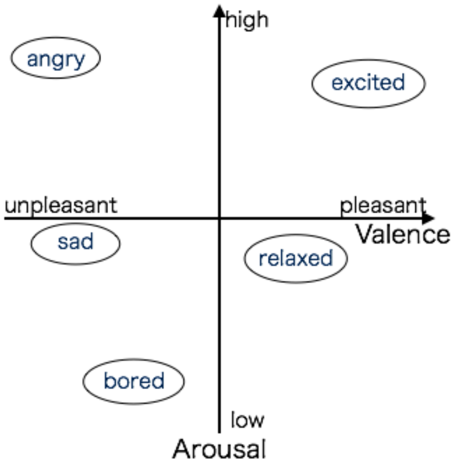

Figure 1: Typical emotions in valence and arousal space.

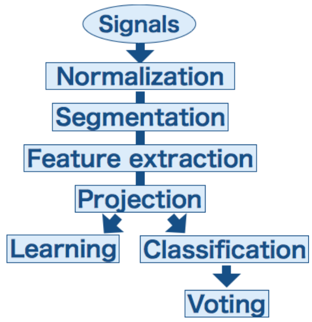

Figure 2: Overview of process to estimate arousal.

comics. Recognizing emotion while reading is more difficult than while watching videos because videos have more stimuli such as a big and surprising sound.

Emotions can be described by two dimensions called emotional valence and arousal (simply called valence and arousal hereafter) [11]. Valence ranges from unpleasant to pleasant. Arousal ranges from low to high. Figure 1 shows some typical emotions are represented in the two-dimensional space. For example, if a reader feels very unpleasant and neutral arousal, the named emotion is sad. As a first step towards capturing various emotions in the two dimensional space, we focus in this paper to estimate the arousal. To achieve this goal, the method proposed in this paper takes as input physiological signals measured by the E4 wristband produced by Empatica, and a stationary eye tracker called RED 250 produced by SMI. This setup can easily be applied in every day life environment. The proposed method extracts features from the signals and applies the support vector machine (SVM) to estimate the arousal. In this paper, we examine the accuracy of the proposed method.

2. RELATED WORK

Many researches estimating users' emotions have been done. Emotions influence users' expression. An approach for the emotion estimation is to analyze the voice and facial expressions [7, 12]. Analyzing voice expression is useless for our research, because while reading comics the readers do not speak. From the facial images captured by a camera, the facial expression can be recognized by software such as Affdex produced by Affectiva. It identifies key landmarks on the face such as the corners of the eyes and extracts regions of the parts on the face. After that, it analyzes the features based on pixel color, texture and gradient with a machine learning method. The probability of each facial expression is outputted as the range from 0 to 1. For the facial expression, the categorization of basic six emotions (e.g., anger, joy, sad, disgust, surprise, fear) proposed by Ekman et al. is used [4]. Zhao et al. proposed a recommendation system of videos based on the facial expression [12]. The method recommends the video if the facial expression changes enough to consider the video as an enjoyable video. We have tried to use facial expressions while reading comics to know emotions of readers. However, it has turned out that the facial expression is not so informative for comics because readers do not change their facial expressions so much.

It is known that the emotions influence physiological signals such as the heart rate, breath pattern and perspiration. Researchers have tried to estimate the users' emotions by analyzing the physiological signals. Haag et al. [6] have applied a neural network technique to features extracted from electromyography, electrodermal activity, skin temperature, blood volume pulse, electrocardiogram, and respiration in order to estimate the emotions. A database with the facial expressions and the physical signals while watching videos is available [9].

Although many researchers have tried to estimate the emotions, most existing work has employed uncomfortable devices. Furthermore, little work has been done to analyze the behavior of reading comics.

3. METHOD

Figure 2 shows the overview of the process to estimate the arousal. First, readers' physiological signals are measured. Then, the signals are normalized and segmented. Features are extracted from each segment and projected onto one dimension. The segment is classified into two classes: high arousal or low arousal. For testing, the whole recording is classified by voting. We describe each step from Sect. 3.1 to 3.5.

3.1 Signals

We measured physiological signals while reading comics. Figure 3 shows the measurement. We employ electrodermal activity (EDA), blood volume pulse (BVP), heart rate (HR), skin temperature (ST) and pupil diameter (PD).

The EDA, BVP, HR and ST are measured by a wristband called E4 wristband. When a human feels arousal, the brain sends signals to the skin and this causes the skin sweating. The E4 wristband measures the electrical conductance which is affected by the amount of sweat. The conductance variation is called EDA and also called galvanic skin response. The EDA can be decomposed in tonic skin conductance level (SCL) and phasic skin conductance response (SCR). The process is done by the software package Ledalab [1].

The PD is measured by a stationary eye tracker attached on a screen displaying digital comics. An advantage using the eye tracker is that the readers do not have to wear any

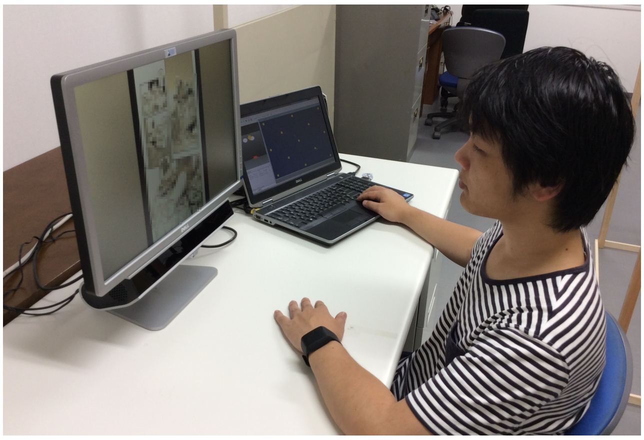

Figure 3: Reading comics while measuring physiological signals by eye tracker and wristband.

devices for measuring PDs.

The outputs of the E4 wristband and RED250 include the timestamp of the recordings. The signals are synchronized based on the timestamp.

3.2 Normalization

The scales and values of the each physiological signals differ. In order to reduce influence of the difference to the classifier, all the signal values are normalized. A value d of a signal is normalized to d_n as follows:

$$d_n = \frac{d - d_{\min}}{d_{\max} - d_{\min}} \quad (1)$$

where $d_{\min}$ is the smallest value and $d_{\max}$ is the largest value of each signal.

3.3 Segmentation

The emotions influence physiological signals within a certain period time usually from seconds to minutes. In order to segment and extract the influenced parts, the signals are processed with a sliding window technique. The size of the window is T_w seconds. The window is slid with the step of T_s seconds. In our experiment, T_w is 30 seconds and T_s is 10 seconds.

3.4 Feature extraction and projection

The averages and variances are calculated as features from each segment. Overall twelve features are extracted from the six signals (SCL, SCR, BVP, HR, ST, PD). Then, we applied linear discriminant analysis (LDA) [3] to reduce the feature dimensionality. The LDA projects the features onto one dimension. This projection removes noisy and meaningless features. For the implementation of LDA, we use a library called scikit-learn [10].

3.5 Learning, classification and voting

Each segment is classified into two classes: high arousal or low arousal. SVM is used as a classfier. The parameters for the SVM are computed by a grid search. For the implementation of SVM and grid search, we also use scikit-learn.

For testing, features of test data are calculated by the same process with the training data. From the classification result of each segment, a majority vote is applied. If the number of segments classified as high arousal is bigger than or equal to half of the total number of segments, the whole recording is classified as high arousal.

Table 1: Accuracy of arousal estimation.

	P1	P2	P3	P4	P5
Accuracy [%]	18.2	90.9	90.9	63.6	27.3
BR [%]	54.5	81.8	54.5	54.5	72.7

4. EXPERIMENT

We conducted an experiment to examine the accuracy of our method to estimate the arousal.

4.1 Experimental setup

We prepared 11 comics and asked 5 participants to read the first chapters displayed on a screen. While reading the comics, we measured the participants' physiological signals. The procedure of the recording was as follows:

1. calibrate the eye tracker
2. read the first chapter of a comic
3. answer the arousal while reading on 10 points scale (from low to high)
4. repeat 1, 2 and 3

The number of pages of the chapters were 60, 24, 61, 65, 41, 59, 50, 19, 10, 14, 39. We randomized the reading order of the comics for each participant. The participants are Japanese male students of the engineering department of university. The ages of the participants range between 20 and 25. The participant no.3 wore contact lenses and the other participants wear neither glasses nor contact lenses. Going to the next page and coming back to the previous pages were done by pushing the buttons of the keyboard put in the front of the screen. The light condition was constant. The sampling rates of the EDA, BVP, HR, ST and PD measuring were 4, 64, 1, 4 and 250 Hz, respectively.

For each participant, we calculated the average of the arousal for all chapters. If the arousal value of a comic is equal or more than the average, we labeled the signals of the comic as high arousal. If the value is less than the average, we labeled the signals as low arousal.

Learning has been done by assuming that the same arousal is obtained throughout a chapter. To be precise, all segments are labeled with the same arousal (low or high) in each chapter.

We applied leave one out cross validation with user dependent training. In other words, the signals for one of 11 comics are chosen as test data and the signals for the other 10 comics are chosen as the training data. We applied our method to the data and repeated this 11 times. We took the average of the accurate rates.

4.2 Result

Table 1 shows the accuracy of our arousal estimation. The first row shows the accuracy of the estimation for each participant. The second row shows the base rate (BR) of the estimation. The BR is the rate if a classifier always estimates

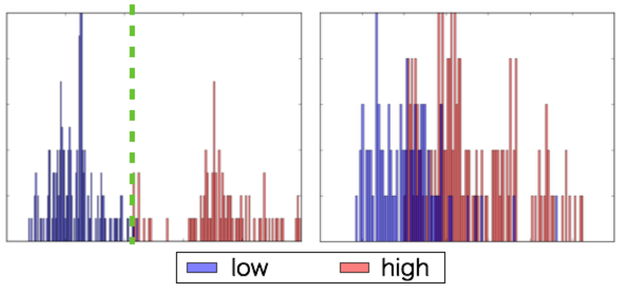

Figure 4: Two examples of histograms of values of the features projected by LDA. The left one shows a good example of distinguishable values in correct estimation. The right one shows, on the other hand an example where most segments of the test data were not classified into the correct classes.

the class according to the majority class in the training data. P1–P5 represent participants.

The results for participants no.2, no.3, and no.4 were over the BRs as shown in Table 1. On the other hand, the results for participants no.1 and no.5 were under the BRs.

Figure 4 shows histograms of values of the features projected by LDA. The x-axis shows values of the projected features. The blue and red bars show the frequency of the values for low and high arousal, respectively. The left one shows the values for the participant no.3 and the right one shows the values for the participant no.5. For the case shown in the left of Fig. 4, low and high are not so overlapped and can be easily distinguished by the green dot line. This means LDA works to separate the features for the participant no.3. On the other hand, it is difficult to distinguish low and high for the right histogram because they are overlapped.

One of the reasons for the failure is as follows. We assume that all segments are with the same arousal. However, user's arousal level can vary depending on the contents even in a single chapter. For some cases, our assumption did not hold. For example, even if the readers answered as they felt high arousal, they might feel low arousal partly. Each segment should be labeled with the emotions elicited at that moment. A difficult problem of labeling each segment is that it is needed to interrupt reading to answer emotions frequently, because it is difficult for readers to recall detailed changes of emotions if we ask infrequently. Interrupting reading disturbs the readers from empathizing and influences the physiological signals.

Moreover, readers' movements influence the signals. We asked the participants not to move as much as possible. However, human moves unconsciously. For example, the participant no.1 taps his finger during some parts of the recording. It is assumed the tapping strongly influenced the signals of the E4 wristband. We consider that the noise from the tapping decreases the accuracy.

5. CONCLUSION

Estimating emotions while reading enables new services such as comic recommendation. Emotions can be described by valence and arousal. In this paper, we propose a method for estimating arousal while reading comics based on analysis of physiological signals measured by E4 wristband and RED250. Our experiment shows that the physiological signals are informative to know the arousal for some readers. However, the arousal for the others are not estimated accurately because of the change of arousal level in a chapter.

Future work includes labeling the physiological signals more precisely and decreasing the influence of readers' movements. Moreover, in order to improve the accuracy, we are planing to employ other sensors and different machine learning techniques. Estimation of valence is also an important study.

ACKNOWLEDGMENTS

This work has supported in part by the JST CREST (25240028), JSPS Kakenhi (15K12172, 16758390).

REFERENCES

[1] M. Benedek and C. Kaernbach. Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*, 47(4):647–658, 2010.

[2] D. O. Bos. Eeg-based emotion recognition. *The Influence of Visual and Auditory Stimuli*, pages 1–17, 2006.

[3] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. John Wiley & Sons, 2012.

[4] P. Ekman, W. V. Friesen, M. O'Sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti, et al. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology*, 53(4):712, 1987.

[5] J. J. Gross and R. W. Levenson. Emotion elicitation using films. *Cognition & emotion*, 9(1):87–108, 1995.

[6] A. Haag, S. Goronzy, P. Schaich, and J. Williams. Emotion recognition using bio-sensors: First steps towards an automatic system. In *Tutorial and research workshop on affective dialogue systems*, pages 36–48. Springer, 2004.

[7] A. B. Ingale and D. Chaudhari. Speech emotion recognition. *International Journal of Soft Computing and Engineering (IJSCE)*, 2(1):235–238, 2012.

[8] Comic plus. http://kc.kodansha.co.jp/search.

[9] S. Koelstra, C. Mühl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. Deap: A database for emotion analysis; using physiological signals. *Affective Computing, IEEE Transactions on*, 3(1):18–31, 2012.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[11] J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

[12] S. Zhao, H. Yao, and X. Sun. Video classification and recommendation based on affective analysis of viewers. *Neurocomputing*, 119:101–110, 2013.