XCSF for Prediction on Emotion Induced by Image Based on Dimensional Theory of Emotion

Po-Ming Lee
Institute of Computer Science and Engineering, National Chiao Tung University, Taiwan (R.O.C)
+886 (3)571 2121#59329
pmli@cs.nctu.edu.tw

Yun Teng
College of Computer Science , National Chiao Tung University, Taiwan (R.O.C.)
1001 Ta Hsueh Rd., Hsinchu, Taiwan (R.O.C.)
yunteng.cs98@gmail.com

Tzu-Chien Hsiao
Institute of Computer Science and Engineering and with Institute of Biomedical Engineering, National Chiao Tung University, Taiwan (R.O.C.).
labview@cs.nctu.edu.tw

ABSTRACT
Affective image classification problem is a problem aims on classifying images according to their affective characteristics of inducing human emotions. This paper extends the discrete state classification problem into a continuous function approximation problem by applying the experimental paradigm of dimensional emotion model. The Extended Classifier System for Function Approximation (XCSF) was applied to the problem and the results suggest that it outperforms linear regression (LR) in accomplishing this task. The obtained results also indicate that without using content based features of the images, the effects of individual difference can be relatively small.

Categories and Subject Descriptors
I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods and Search

General Terms
Algorithms, Measurement, Experimentation, Human Factors.

Keywords
Extended Classifier System, Affective Picture, Self-Assessment Manikin

1. INTRODUCTION
People experience emotions by feeling happy, angry and various emotions induced by stimulus and events in their daily life, it is crucial to understand why, what, and how the emotional system of human works. During past decades, Lang et al. (1995) reported an empirical study on the effects of emotion characteristics of video tapes that affect cognitive capacity and memory of subjects [15]. Subsequently, Bolls et al. (2001) revealed that subjects tend to remember stimulus that elicit negative emotions more, than the stimulus that elicit positive emotions [1]. Later researchers focused on the use of the stimulus on attracting the attention of subjects, and to make subjects remember more on the presented items. For example, the use of affective image for eliciting emotions in print advertisement [8], the study on how emotions induced by affective features affect candidate evaluation in political advertisement [7], and the use of affective stimulus for attracting the attention of people in web browsing, which a guideline of extracting affective features in web pages was also reported [14]. Lee et al. (2010) furthermore demonstrated the effect of emotion on decision making process of human and how can such effect be detected through heart rate information collected from healthy subjects [18].

The main stream of emotion researches can be categorized into four categories: 1) what emotion is elicited by certain stimuli; 2) the role of distinct stimulus plays in emotion elicitation; 3) the understanding on how human emotion is induced and 4) why are the human emotions being induced. This paper falls into the second one, that is, the role of distinctive stimuli plays in emotion elicitation. We aimed on mining the affective characteristics contained in images that are responsible for emotion elicitation. While it is human nature to pursue happiness and avoid pain, the results obtained in this study may be transferred to diverse fields that relate to "human", for example, media studies and advertisement.

1.1 Literature Study on Affective Image Classification
To predict emotions of subject induced by an image, in 2005, Mikels et al. firstly categorized images in International Affective Picture System (IAPS) (details in section 2.2.1) into different categories to identify images that are especially excellent in inducing emotions of subjects [21]. Later, a pioneering study [29] on affective image classification contributed by Wu et al.. The study applied the Support Vector Machine (SVM) on identifying the relationships between visual features extracted from images, and the semantic differential features, that is, terms that given to subjects to describe the onset image, such as beautiful-ugly, dynamic-static, and tense-relaxed. The accuracy rate obtained in the study was relatively high (i.e. 80%), however, only one participant was involved in the experiment, and the emotion was implicitly estimated through semantic differential terms. To demonstrate the feasibility in affective image classification, subsequently, experiments with larger sample size (around 15 to 20 people) was conducted in [19, 30]; in those studies, emotions were explicitly defined as discrete emotional states, such as happy, surprising, sad, and angry. Further examinations on various features in affective image classification task was reported by Machajdik in [20]; however, the obtained accuracy rates remain relatively low (around 65%) in the between subject analysis in [19, 20, 30]. Latest findings in [2, 3] highlighted the drawback on using discrete emotion models, in which definition on emotions

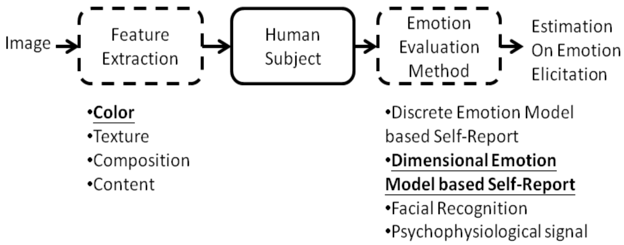

Figure 1 The affective image classification problem as a system identification task

using "terms" may be vague and inaccurate for the subjects, and the use of discrete emotion model is generally application dependent, which may bias the collected dataset, and affect the classification performance. Furthermore, the use of discrete definitions also makes the experiment results hard to reproduce and hard to compare internationally. Hence, we argue that the affective image classification studies should be conducted based on dimensional emotion model (details see section 2.1) to reduce the difficulties in reproducing comparable results.

1.2 Research Objectives

To clarify the objectives, the affective image classification problem is formatted into a system identification task (see Figure 1), the aim of the problem is to identify how human subjects interpret the affective characteristics of a given image; for example, to identify the human subject response by discovering rules, or training intelligent systems to predict the response (currently most of the works aimed on the later approach).

To evaluate the emotion elicitation of a subject, despite numerous approaches are available, for example, self-report [4], facial expression [9], and psychophysiological data [13], we selected self-report as our measurement tool in this study since taking self-report as a ground truth is considered to be more meaningful in the proposed problem and also the related future applications.

This study conducts an experimental study on affective image classification by adopting dimensional emotion model instead of applying discrete emotion models that were typically used in previous studies [20]. A pictorial based self-report named Self-Assessment Manikin (SAM) (details see section 2.2.2), which is commonly used in the field of emotion research, was used in this study to estimate the emotion elicitation of subjects in the perspectives of dimensional emotion model. The use of the experimental paradigm of dimensional emotion model, on the other hand, extends the traditional discrete emotion classification task into a continuous function approximation task, hence, in this study, the performance of classifiers were judged by Root Mean Square Error (RMSE) for two-dimension affective space prediction, and Mean Absolute Error (MAE) for one-dimension affective space prediction, instead of accuracy rate.

The objective of the study is to upgrade the performance of affective image classification; however, as mentioned in last section, the results of previous studies, that obtained from experimental designs based on discrete emotional model, are hard to compare, and even incomparable in our case (because of the shift of performance criteria), hence, in this study, we just examine the performance obtained from the described task, and hope to provide a baseline for future study instead.

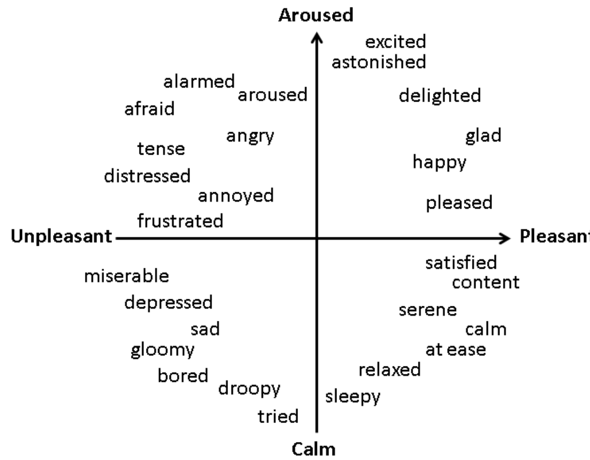

Figure 2 Defining emotions in a two-dimension affective space

1.3 Paper Contributions

This paper provides three major contributions. First, through the analysis of the prediction results of the Extended Classifier System for Function Approximation (XCSF) (details see section 2.4, and section 4), the paper demonstrates the capability of XCSF on function approximation task in affective image classification problem. Second, by demonstrating low error in regard to RMSE/MAE in the function approximation task in affective image classification problem, this paper provides support on adopting dimensional emotion model in affective image classification problem. Third, by carefully following the protocols and standard instruments typically used in the field of psychology for emotion induction (details see section 3) [17], the results can be easily reproduced, and hence the performance achieved by XCSF in this study may also be regarded as a baseline performance in affective image classification problem with dimensional emotion model adopted.

1.4 Paper Organization

The remainder of this paper is structured as follows: Section 2 provides a review on latest findings, theory and methodology revealed in the area of psychology and neuroscience, and also presents XCSF and the adopted data analysis process; Section 3 presents the experiment conducted and the collected dataset; lastly, obtained results and further discussion are provided in Section 4.

2. METHODS AND MATERIALS

2.1 Dimensional Theory of Emotion

To clarify the definition of "emotion", this section provides further details on emotion model adopted in this paper. Although there is a tendency for researchers to intuitively define a set of discrete, basic emotions, such as: happy, surprising, sad, angry, etc. [22], recently the dimensional theory in replacement of traditional emotion models has been proposed and shown superior in various studies [2, 3].

Dimensional theory defines emotion by a two dimensional affective space: valence and arousal. Valence represents whether the emotion experiencing is pleasant, and arousal represents how arousing the experiencing emotion is. The definition of emotion in this study is illustrated in Figure 2. The reason of why a stimuli is

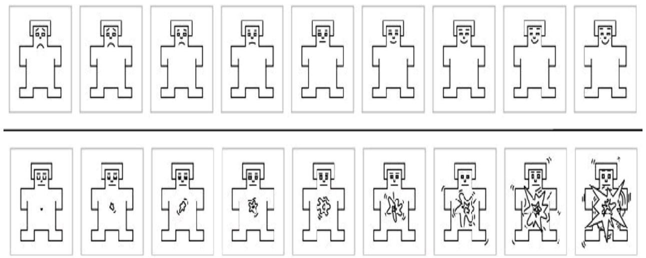

Figure 3 Self-Assessment Manikin (SAM) used in this study, in which the upper row represents valence and the lower row represents arousal

emotionally relevant and how a stimuli is emotionally relevant[1] can be considered as a result of evolutionary process, because stimuli that stimulate positive emotions are usually found related to food or sex, whereas stimuli that stimulate negative emotions are usually found related to danger or death, namely, the need of survival [16]. The dimensional theory has attracted substantial attention in the field of psychology since proposed, and has been adopted in numerous studies [1, 2, 7].

2.2 International Affective Picture System (IAPS) and SAM

2.2.1 IAPS

The images used in this study were solely from a public database named IAPS, which is developed by Lang and Bradley [17] and can be easily obtained from through an e-mail application. Image is a type of visual stimulus that commonly used in human emotion study for emotion induction. However, in past decades, due to cultural difference, results obtained from different experiments were incomparable. Subsequently, a standard affective picture system named International Affective Picture System (IAPS) was proposed [17] to help emotion research experimenter in providing comparable experimental results. The IAPS contains various affective pictures selected based on the statistics obtained from experimental results, that proved to be capable in inducing diverse emotions in the affective space. The IAPS has attracted attention since proposal; various experiments, for example, the empirical studies on psychophysiological signals that are related to emotional responses [25], the experiments of the effects of emotion on memory [12], and the experiments for identifying the relationship between motivation and emotion [3], were conducted using IAPS for emotion induction.

2.2.2 SAM

As mentioned in section 1.2, a pictorial based self-report named SAM which realized the dimensional emotion model is used in our experiment to have subjects self-report emotion elicited by the onset of each image.

Recent efforts on emotion studies focused on the means to index human emotion. Although human emotion is used to be hard to measure because of the lack of ground truth, during the development of instruments, researchers utilized verbal assessment to have subjects assess their own emotion responses. However, verbal assessment using phrases to describe emotions is sometimes considered to be vague due to cultural difference and individual difference on the definition of phrases. Another disadvantage is that the assessment using phrases to describe emotions inherently adopts the paradigm of discrete emotions.

Hence, to manage the disadvantages and to realize a self-report instrument complies with dimensional emotion model, a non-verbal pictorial assessment named Self-Assessment Manikin (SAM) was proposed by Bradley and Lang in which the format of the assessment is provided in Figure 3. Since proposal, SAM has been successfully used in diverse theoretical studies and applications [4], hence, the SAM is used in our experiment to provide internationally comparable self-report data.

2.3 HSV Model

This study adopt the approach of feature extraction similar to the former studies [20, 30], in which only basic features based on colors were extraction from the image (HSV model, in our case) instead of applying content based analysis, to eliminate the individual difference. Texture information was not used in this study because of the documentary-style natural of the IAPS images; images in the IAPS hold similar texture properties, and the related features extracted from IAPS images was reported useless in [20].

The HSV model is a cylindrical-coordinate representation commonly used in the area of computer graphics in replacement of RGB color model to obtain more intuitive values. In the HSV model, H represents hue, S represents saturation, and V represents value. Ordinarily, images stored in electronic devices such as personal computer are represented by a $\mathrm{M} \times \mathrm{N}$ matrix, in which the color of each element is displayed using RGB color model. The RGB model is a model consists of three coordinates as following: R represents red values, G represents green values and B represents blue values; red, green and blue are mixed together in a cube. For affective features analysis, features extracted from HSV model provide a more perceptually relevant representation on images.

2.3.1 Hue

Hue is simply the attribute represents visual sensation on various colors similar to red, green, blue, or combinations of them. The value of hue is in the interval of 0° and 360° (normalized to interval [0, 1] in this study). The transformation from RGB to H is demonstrated as following: Firstly, normalizes R, G, and B of the target element into the interval [0, 1]. Secondly, calculates M, m and C from the normalized R, G, B.

$$\mathrm{M} = \max(\mathrm{R}, \mathrm{G}, \mathrm{B});\ \mathrm{m} = \min(\mathrm{R}, \mathrm{G}, \mathrm{B});\ \mathrm{C} = \mathrm{M} - \mathrm{m} \qquad (1)$$

Thirdly, calculates H' and H.

$$\mathrm{H}' = \begin{cases} 0, & \text{if } \mathrm{C} = 0 \\ \frac{\mathrm{G}-\mathrm{B}}{\mathrm{C}} \bmod 6, & \text{if } \mathrm{M} = \mathrm{R} \\ \frac{\mathrm{B}-\mathrm{R}}{\mathrm{C}} + 2, & \text{if } \mathrm{M} = \mathrm{G} \\ \frac{\mathrm{R}-\mathrm{G}}{\mathrm{C}} + 4, & \text{if } \mathrm{M} = \mathrm{B} \end{cases} \qquad (2)$$

$$\mathrm{H} = 60^\circ \times \mathrm{H}' \qquad (3)$$

[1] The umbrella term "emotionally relevant stimuli" can be simply understood as stimuli that are able to elicit certain emotions of a person (either positive emotions or negative emotions)

2.3.2 Saturation

Saturation represents the level of colorfulness relative to its own brightness. The value of saturation is in the interval [0, 1].

$$S = \begin{cases} 0, & \text{if } C = 0 \\ \frac{C}{M}, & \text{o.w.} \end{cases} \quad (4)$$

2.3.3 Value (Brightness)

The Value (brightness) represents the brightness level relative to the brightness of a similarly illuminated white, defined as the largest component of the RGB color of an element (i.e, M, $0 \leq M \leq 1$) to form a hexagonal pyramid out of the RGB cube by projecting all three primary colors and the secondary colors such as cyan, yellow, and magenta into the new plane.

2.4 XCSF

Instead of utilizing previous approaches such as SVM, Naïve Bayes Classifier, and Decision Tree (DT), the paper utilizes a piece-wise linear approximation method named Extended Classifier System for Function Approximation (XCSF) for the function approximation task of affective image classification problem described in section 1.2 to minimize the RMSE/MAE by coping with possible non-linearity contained in the collected dataset. The linear regression (LR) [11] was also applied to the task to compare the performance achieved by XCSF.

2.4.1 XCS

This section provides a brief description of XCS. The XCS is a machine learning system based on Michigan-Style Classifier Systems (CSs). Since the 1990s, XCS has attracted considerable attention in the field of CSs because of its theoretical advances and its applicability in practice [23, 24]. The XCS is a rule-based online learning algorithm, which can extract knowledge from a previous unknown dataset in an iterative manner. The XCS can also be regarded as a system that manages a set of classifiers that are represented in the traditional production system form of "IF state THEN action". By integrating the Genetic Algorithm (GA) component (also named rule discovery component in XCS), the set of classifiers can evolve occasionally, and search for a set of classifiers that yields the maximal generality and accuracy. The original XCS was designed to perform on datasets with discrete inputs and a discrete output.

2.4.2 XCSF

In 2002, XCSF, as a version of XCS used for function approximation, allowing both real value inputs and real value outputs (furthermore, the version of XCSF implemented in [26] allows multiple outputs), was proposed [27]. The input was extended to real value by using rotating hyperrectangle and rotating hyperellipsoid for condition representation [6, 28]. Instead of selecting a discrete value as output according to fitness-weighted prediction value, XCSF classifiers directly map the desire output using the prediction value produced by a linear approximation (i.e, $h(\vec{x}) = \vec{\omega}\,\vec{x}$ where $\vec{x}$ represents the input vector and $\vec{\omega}$ represents weight vector). Each classifier in XCSF updates its weight vector using Recursive Least Squares (RLS) method [6]. For using RLS, each classifier manage by XCSF updates its weight vector using

$$\vec{\omega} \leftarrow \vec{\omega} + \vec{k}\left[y_t - \left(\vec{x^*} - \vec{m^*}\right)^T \vec{\omega}\right] \quad (5)$$

where y_t represents target output, and $\vec{k}$ represents the gain vector computed by

$$\vec{k} = \frac{V^T\left(\vec{x^*} - \vec{m^*}\right)}{\lambda + \left(\vec{x^*} - \vec{m^*}\right)^T V^T \left(\vec{x^*} - \vec{m^*}\right)}. \quad (6)$$

The λ (usually $0 \leq \lambda \leq 1$) used in Formula 6 and 7 represents a forget rate for RLS, the lower the value is the higher the forget rate, in which λ set to 1.0 represents for infinite memory (mostly used in time invariant problems). The matrix V hold by each classifier updates recursively using

$$V^T = \lambda^{-1}\left[I - \vec{k}\left(\vec{x^*} - \vec{m^*}\right)^T\right] V^T. \quad (7)$$

The fitness value used for GA in XCSF is based on relative classifier accuracy calculates from system error [5]. For further detail, sufficient information on XCS can be found in Butz's algorithmic description of XCS [5], and recent advances in XCSF [6, 27, 28]. In summary, XCSF can be understood as a manager manages a set of classifiers in which the each classifier maps from a subspace in the feature space to the landscape function output using a linear fitting method.

3. EXPERIMENT DESIGN

The dataset used in this study for performing affective image classification (formation as a system identification task, see Figure 1) was collected from a human subject experiment, in which images selected from IAPS were used as the system input and SAM was used as the system output.

3.1 Task Description

The entire experiment conducted in this study comply the IAPS protocol of emotion induction described in [17] to guarantee the effectiveness of the emotion induction procedure, and the clarity of the experimental design for reproduction. During the experiment, the subjects were requested to look at a screen which sequentially presents images, and rate the presented images correspondingly in regard to the induced emotions, that is, valence and arousal, using computer-based SAM through mouse. The period length of the experiment was 10 minutes for each subject. Each trial (presentation of an image) was started with a presentation of an image, displayed for 6 seconds, then a presentation of SAM on the screen for subject to manually rate the affective characteristics (self-report the induced emotion) of the presented image, followed by a 15 seconds delay to ensure the emotional status of subject return to baseline before the start of next trial and at the same time to create a reasonable length to keep subjects involved in the experiment.

3.2 Environment Setting

During the experiment all subjects were presented with 20 non-repeated images selected from IAPS according to IAPS picture set selection protocol described in [17], which states the constraint on the number of images used in a single experiment, and the constraint on the distribution of the emotions that are expected to be induced by selected images, etc.. The exact id of the used picture are: 1120, 1310, 1390, 1710, 1720, 2160, 2220, 2520, 2530, 2540, 3160, 3220, 3250, 4300, 4460, 4470, 4660, 4750, 5950, 8160, 8200, 9250 (those images can be found in IAPS picture set [17] through the numbers listed above).

The order of image presentation was randomized to eliminate effects due to the sequence. The images were presented using a general PC with 32-inch (81.28 centimeters) monitor. Subjects were sat in a comfortable bed at a distance of approximately 1.5 meters away from monitor in an EMI shielding room (Acoustic Inc. US) in which eliminates most of noise interferences and electrical noises; the CO2 concentration of the environment was

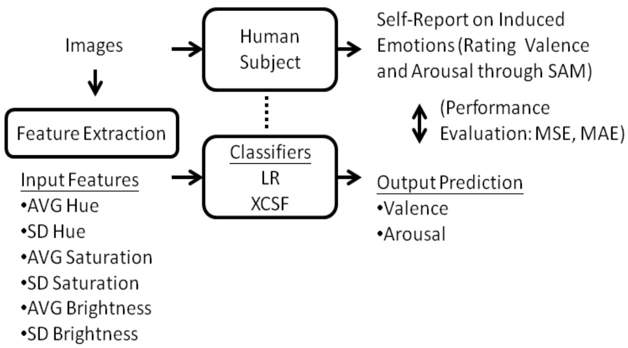

Figure 4 The workflow of the built prediction model

monitored during the entire experiment to guarantee reasonable CO2 concentration (500 ppm ~ 1,300 ppm) to keep subjects sustained its attention during the experiment. In addition, during the experiment the psychophysiological data of subject was also collected in conjuncture with SAM in anticipation of future analysis. The Electrocardiogram (ECG) was recorded from three Ag/AgCl reference electrodes, attached via the lead-2 placement with signals amplified by an amplifier, and also Skin Conductance Response (SCR), were acquired by DAQCard (National Instruments, US) with sampling rate of 1 kHz.

3.3 Subject Selection

There were 16 university students participated in the study (15 subjects is the minimum sample size required in the field of affective image classification studies [29, 30]), ranging in age between 20 and 28 (mean = 23.44, standard deviation = 2.19; 10 men, 6 women). All subjects reported they were healthy, with no history of brain injury, cardiovascular problems, and had normal or corrected-to-normal vision and normal range of finger movement. The experiment and the manner of using data obtained from human subject were approved (Protocol No: 100-014-E) by the Institution Review Board (IRB) of the National Taiwan University Hospital Hsin-chu Branch.

3.4 Collected Dataset

3.4.1 Dataset Description

The collected dataset contains 20 images (1024x768 JPEG) used in the experiment, and the image affective ratings rated by 16 subjects through SAM. The experiment totally acquired 320 rows (actually, 318 rows, while two rows were excluded due to machine mal functioning) of raw data (images, and the affective ratings of the images, 20 rows for each subject). Figure 5 presents the distribution of the ratings selected by subjects on all images; it is observed that most subjects were aroused with either unpleasant feelings or pleasant feelings by the displayed images, no obvious skewed was observed in the distribution of valence (histogram was examined but not shown).

3.4.2 Preprocessing and Data Analysis

The workflow of the preprocessing and model building are provided in Figure 4; the preprocessing of the image data was based on HSV model (details see section 2.3) without applying content based analysis, 6 features were used for model building in this study, including: average hue, standard deviation of hue, average saturation, standard deviation of saturation, average brightness, and standard deviation of brightness. The model was built to predict the induced emotion rated by subjects in terms of valence and arousal through SAM. The prediction of valence and arousal can be real number herein according to the definition of valence and arousal in the dimensional theory of emotion [2]. To avoid over-fitting problem, a Leave-One-Out-Cross-Validation (LOOCV) on leaving one sample at each time for testing set and the remain samples for training set, which is the standard practice for analyzing limited dataset, was used for building the model. The 10-Fold Cross Validation (10-Fold CV) was also used in our study, however, the use of 10-Fold CV substantially decreased the performance achieved by used classifiers (results not shown).

Table 1 The performance achieved by distinct classifiers

Prediction Results	Affective Dimension				
	Valence		Arousal		(Valence, Arousal)
Method	MAE	SD	MAE	SD	RMSE
uniRand	2.569541	(N/A)	2.530387	(N/A)	(N/A)
largCount	1.613781	(N/A)	1.480539	(N/A)	(N/A)
LR	1.482981	1.021157	1.481865	1.070608	2.564612
XCSF	0.970546	0.747286	1.460724	1.029923	2.165585

For details on the setting of LR and XCSF for building the models, the LR analysis was done by using the Weka implementation of data mining tools [10], in which Akaike criterion was used for model selection and M5's method was used for attribute selection; all the co-linear attributes were excluded.

The XCSF used in this study was adopted from the Java implementation version on XCSF contributed by Stalph and Butz (2009) [26]. For parameters setting, α = 1.0; β = 0.1; δ = 0.1; λ = 1.0; θ_{ga} = 50; ε_0 = 0.5; δ_{rls} = 1000; θ_{del} = 20; χ = 1.0; μ = 1.0; θ_{sub} = 20; GA subsumption was turned on. Because during the entire training the number of classifiers was quickly converged to nearly 5,400, the maximal population size N was set to 6,400 to maximize the performance of XCSF. To examine the performance of the system, ε_0 was set to various values, but relatively small effects on the performance were found. During model training, the XCSF was sequentially presented with 20,000 instances randomly selected from the training dataset.

4. RESULTS AND DISCUSSION

The performance in regard to RMSE/MAE and the standard deviation of MAEs (represents by SD) achieved by LR and XCSF are provided in Table 1. The manner of calculating RMSE and MAE are provided as following:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left[(V_i - VP_i)^2 + (A_i - AP_i)^2\right]}$$

$$\mathrm{MAE} = \frac{1}{N-1}\sum_{i=1}^{N}|V_i - VP_i| \text{ or } \frac{1}{N-1}\sum_{i=1}^{N}|A_i - AP_i|$$

in which *N* represents sample size, V_i and A_i represents the valence and arousal corresponds to the *i*-th sample; VP_i and AP_i represents the system prediction on the valence and arousal

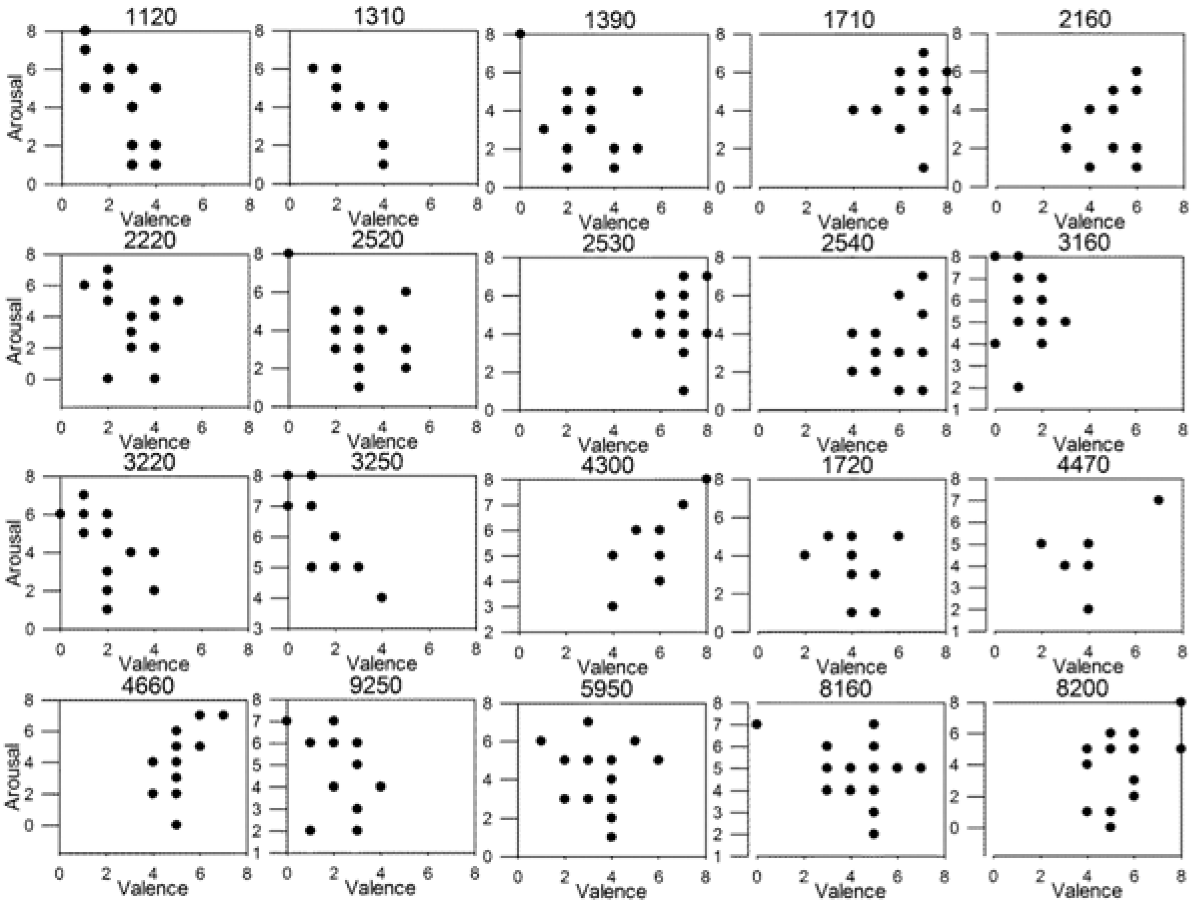

Figure 5 The distribution of the induced emotion of subjects on all images

corresponds to the *i*-th sample. The RMSE is always adopted when a classifier is used for predicting valence and arousal in pairs, and the MAEs are always adopted when a classifier is used for predicting valence and arousal separately.

While the emotional ratings are not uniformly distributed, the MAE of prediction can be artificially underestimated; hence, two models, 1) uniRand: making predictions in a uniformly random manner, and 2) largCount: making constant predictions based on the weighted-average valence, and weighted-average arousal, based on the ratings, in which average value of valance was nearly 3.931 and average value of arousal was nearly 4.349, were introduced to compare the MAE achieved by LR and XCSF. The performance of distinct classifiers is provided in Table 1. The performance of LR on predicting valence in regard to MAE is 1.483, which is relatively low while the uniRand achieved only 2.570 and largCount achieved 1.614. In addition, the MAE value achieved by XCSF on predicting valence decreased the MAE value achieved by LR from 1.483±1.02 to 0.971±0.747 (-0.512), demonstrates the capability of XCSF on mapping functions that possibly contain non-linearity by managing a set of linear classifiers. The MAE achieved by XCSF was small and the standard deviation of the MAE is tolerable.

To further examine the performance of XCSF on this task, the performance of classifiers on predicting valence values in terms of MAE are illustrated in Figure 6, in which x-axis represents valence and y-axis represents MAE. The MAE on each valence is represented by four bars: the MAE achieved by uniRand, largCount, XCSF, and LR, respectively. The MAE achieved by XCSF is smaller than the MAE of LR, uniRand and largCount at most of the emotional ratings; the MAE of XCSF is only larger than uniRand at the rating with the largest count. A skew on MAE was observed for those ratings that represent for “being pleasant”, that is, 5~8, possibly due to the sample size of the rating, while in Table 1, the numbers of samples of valence equals to 5, 6, 7 are larger than the numbers of samples of valence equals to 0 and 1.

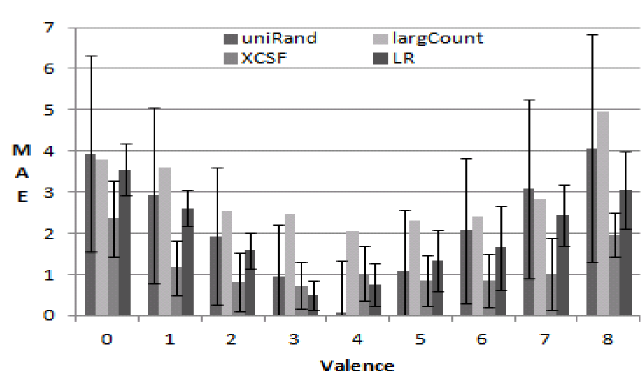

Figure 6 The MAE of XCSF on each valence value. The standard deviation of MAE made by XCSF is nearly 0.54~0.91.

Conversely, the MAE of XCSF in valence 0 and 8 are also high. The finding suggests that insufficient on sample size of a class may lead to low performance of XCSF on approximating the corresponding output value even the training instances were selected from the training dataset randomly during the XCSF iterative training process. However, the MAE of XCSF at valence value equals to 4.0 is not the lowest, indicating that the larger sample size may only guarantee the efficacy of XCSF in function

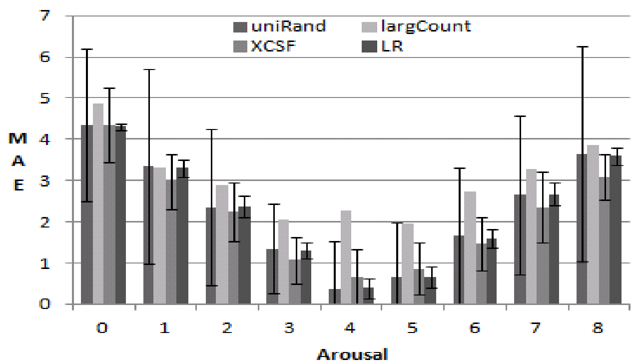

Figure 7 The MAE of XCSF on each arousal value. The standard deviation of MAE made by XCSF is nearly 0.27~0.91

approximation but not eliminating all existing errors. We observed that approximately 30 samples (which is nearly 10% of the number of raw data), is sufficient for XCSF to approximate a valence value in the collected dataset. Otherwise, for example, the MAEs of XCSF on predicting the samples that valence equals to 0 and 8 are relatively high, in which the numbers of samples were small. To explain the extreme cases happened in valence 0 and valence 8 in the psychological perspective, we've noticed that some of the subjects reported that they tend to select ratings in the middle of the scale rather than selecting those ratings represent for extreme cases. Such phenomenon may cause the non-linearity of the distances between the levels of valence and arousal. Further clarification is required for this issue; a well designed transformation may be adequate.

Figure 7 also presents similar phenomenon. The MAEs achieved by XCSF in each level of arousal mostly outperform the MAEs of LR. However, in general, the MAEs achieved by XCSF at each level were increased, and the decreases in error are not sufficiently significant. The results provide that MAE of LR and XCSF did not outperform largCount in the prediction on arousal. Such observation indicates that the prediction model of arousal built by LR and XCSF did not adequately identify the problem structure, possibly due to the ineffectiveness of SAM on estimating subjects' arousal, while some of subjects reported that during the experiment the definition of "being aroused" can be easily confused with the definition of the tendency of valence. The confusion was possibly caused by the cultural difference, but similar results were not highlighted previously in the main stream of the research of SAM.

To further identify the discovered knowledge, the prediction models built by LR are provided in Equation (8) and Equation (9).

$$Valence = -2.3147 * Avg\_Saturation + 4.6681 * Avg\_Brightness + 12.8186 * SD\_Brightness - 0.3798 \qquad (8)$$

$$Arousal = 3.4657 * SD\_Saturation - 3.3625 * SD\_Brightness + 4.3804 \qquad (9)$$

From Equation (8), the saturation of an image tends to lower the valence, whereas the brightness tends to enhance the pleasant feelings, and the effect of standard deviation of brightness on valence is even more substantial. On the other hand, the affective characteristics of an image making people feel aroused, is negative correlated with the standard deviation of brightness, whereas the standard deviation of saturation increases the effect.

Gender information was not used for model building because the effect of gender is eliminated by the use of IAPS protocol, the use of gender information did not substantially improve the performance during our study as well. By contrast to the gender information, within subject analysis was also applied to the dataset using LR and XCSF. The result indicates that without using the content within the image, the effects of individual difference is relatively small (not shown here).

In regard to the concerns of the application of this study, the RGB model is device-dependent, due to the color elements (such as phosphors or dyes) and their response to the individual R, G, and B levels vary from manufacturer to manufacturer, different devices may detect or reproduce a given RGB value distinctively, or even in the same device over time. Such characteristics may cause variations in the proposed experimental results. However, similar problem may also occur in other studies that utilize IAPS, but currently, of our best knowledge, no previous study reports further problem due to it.

5. CONCLUSION

This paper aims on solving affective image classification and has presented the experimental results of adopting the experimental paradigm of dimensional emotion model. The affective image classification problem was reformatted to an affective space function approximation task in objective of minimizing the RMSE of the built prediction model. The XCSF was demonstrated to adequately approximate the output landscape of the affective image classification problem in the collected dataset; the obtained results were compared to the performance of LR.

For suggested future work, applications on indexing affective characteristics of the images on the internet, or adjusting human emotional states to improve the quality of life, for example, to ease people, or to excite people, are possible future work. The future work on using images obtain from other sources, other than IAPS database to validate the experimental results in regard to generality, is suggested.

6. ACKNOWLEDGMENTS

This work was fully supported by Taiwan National Science Council under Grant Number: NSC-100-2220-E-009-041 and NSC-100-2627-E-010-001. This work was also supported in part by the UST-UCSD International Center of Excellence in Advanced Bioengineering sponsored by the Taiwan National Science Council I-RiCE Program under Grant Number: NSC-100-2911-I-009-101.

7. REFERENCES

[1] Bolls, P. D., Lang, A. and Potter, R. F. The Effects of Message Valence and Listener Arousal on Attention, Memory, and Facial Muscular Responses to Radio Advertisements. *Communication Research*, 28, 2001, 627-651.

[2] Bradley, M. M.: 'Emotional memory: a dimensional analysis', in S. H. M. v. Goozen, N. E. v. d. Poll and J. A. Sergeant (Eds.): 'Emotions: Essays on emotion theory' (Lawrence Erlbaum, (1994), pp. 97-134

[3] Bradley, M. M. and Lang, P. J.: ‘Emotion and motivation’, in J. T. Cacioppo, L. G. Tassinary and G. Berntson (Eds.): ‘Handbook of Psychophysiology’ (Cambridge University Press, 2007, 3 edn.), pp. 581-607

[4] Bradley, M. M. and Lang, P. J. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental,* 1994.

[5] Butz, M. and Wilson, S.: ‘An Algorithmic Description of XCS’, in P. Luca Lanzi, W. Stolzmann and S. Wilson (Eds.): ‘Advances in Learning Classifier Systems’ (Springer Berlin / Heidelberg, 2001), pp. 267-274

[6] Butz, M. V., Lanzi, P. L. and Wilson, S. W. Function Approximation With XCS: Hyperellipsoidal Conditions, Recursive Least Squares, and Compaction. *Evolutionary Computation, IEEE Transactions on*, 12, 3 2008, 355-376.

[7] Chang, C. The Impacts of Emotion Elicited By Print Political Advertising on Candidate Evaluation. *Media Psychology*, 3, 2 (2001), 91-118.

[8] Chowdhury, R. M. M. I., Olsen, G. D. and Pracejus, J. W. Affective Responses to Images In Print Advertising: Affect Integration in a Simultaneous Presentation Context. *Journal of Advertising*, 37, 3 2008, 7-18.

[9] Cohen, I., Sebe, N., Garg, A., Chen, L. S. and Huang, T. S. Facial expression recognition from video sequences: temporal and static modeling. *Computer Vision and Image Understanding*, 91, 1–2 2003), 160-187.

[10] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. H. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11, 1 2009, 10-18.

[11] Kaw, A. and Kalu, E. *Numerical Methods with Applications*. autarkaw, 2010.

[12] Kensinger, E. A., Garoff-Eaton, R. J. and Schacter, D. L. Effects of emotion on memory specificity: Memory trade-offs elicited by negative visually arousing stimuli. *Journal of Memory and Language*, 56, 4 2007, 575-591.

[13] Kim, K. H., Bang, S. W. and Kim, S. R. Emotion recognition system using short-term monitoring of physiological signals. *Medical & Biological Engineering & Computing*, 42, 3 2004, 419-427.

[14] Kyung-Sun, K. Effects of emotion control and task on Web searching behavior. *Information Processing & Management*, 44, 1 (2008), 373-385.

[15] Lang, A. and Dhillon, K. The effects of emotional arousal and valence on television viewers' cognitive capacity and memory. *Journal of Broadcasting & Electronic Media*, 39, 3 1995, 313-327.

[16] Lang, P. J. The Emotion Probe - Studies of Motivation and Attention. *American Psychologist*, 50, 5 (May 1995), 372-385.

[17] Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida, Gainesville, FL.

[18] Lee, P.-M., Chang, C.-W. and Hsiao, T.-C. Can human decisions be predicted through heart rate changes? In *Proceedings of the 2010 Second World Congress on Nature and Biologically Inspired Computing (NaBIC)* (15-17 Dec. 2010), 189-193.

[19] Liu, N., Dellandréa, E., Tellez, B. and Chen, L.: ‘Associating Textual Features with Visual Ones to Improve Affective Image Classification’, in S. D’Mello, A. Graesser, B. Schuller and J.-C. Martin (Eds.): ‘Affective Computing and Intelligent Interaction’ (Springer Berlin / Heidelberg, 2011), pp. 195-204

[20] Machajdik, J. and Hanbury, A. Affective image classification using features inspired by psychology and art theory. In *Proceedings of the Proceedings of the international conference on Multimedia* (Firenze, Italy, 2010). ACM, 83-92.

[21] Mikels, J., Fredrickson, B., Larkin, G., Lindberg, C., Maglio, S. and Reuter-Lorenz, P. Emotional category data on images from the international affective picture system. *Behavior Research Methods*, 37, 4 (2005), 626-630.

[22] Ortony, A. and Turner, T. What's basic about basic emotions. *Psychological review* (1990).

[23] Quirin, A., Korczak, J., Butz, M. V. and Goldberg, D. E. Analysis and evaluation of learning classifier systems applied to hyperspectral image classification. In *Proceedings of the Intelligent Systems Design and Applications, 2005. ISDA '05. Proceedings. 5th International Conference on* (8-10 Sept. 2005, 2005), 280-285.

[24] Ravichandran, B., Gandhe, A. and Smith, R. E. XCS for robust automatic target recognition. In *Proceedings of the Proceedings of the 2005 conference on Genetic and evolutionary computation* (Washington DC, USA, 2005). ACM, 1803-1810.

[25] Sanchez-Navarro, J., Martinez-Selva, J., Torrente, G. and Roman, F. Psychophysiological, behavioral, and cognitive indices of the emotional response: A factor-analytic study. *Spanish Journal of Psychology*, 11, 1 (May 2008), 16-25.

[26] Stalph, P. O. & Butz, M. V. (2009), Documentation of JavaXCSF (COBOSLAB Report Y2009N001). Retrieved from University of Wurzburg, Cognitive Bodyspaces: Learning and Behavior website: http://www.uni-wuerzburg.de/fileadmin/ext00209/user_upload/Publications/2009/Stalph09JavaXCSF.pdf

[27] Wilson, S. W. Classifiers that approximate functions. *Natural Computing*, 1, 2 (2002), 211-234.

[28] Wilson, S. W. Get Real! XCS with Continuous-Valued Inputs. *Learning Classifier Systems*, 1813 (2000), 209-219.

[29] Wu, Q., Zhou, C. and Wang, C.: ‘Content-Based Affective Image Classification and Retrieval Using Support Vector Machines’, in J. Tao, T. Tan and R. Picard (Eds.): ‘Affective Computing and Intelligent Interaction’ (Springer Berlin / Heidelberg, 2005), pp. 239-247

[30] Zhang, H., Augilius, E., Honkela, T., Laaksonen, J., Gamper, H. and Alene, H.: ‘Analyzing Emotional Semantics of Abstract Art Using Low-Level Image Features’, in J. Gama, E. Bradley and J. Hollmén (Eds.): ‘Advances in Intelligent Data Analysis X’ (Springer Berlin / Heidelberg, 2011), pp. 413-423