ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
http://www.jiisonline.org
J Intell Inform Syst 2017 March: 23(1): 109~126
http://dx.doi.org/10.13088/jiis.2017.23.1.109

Emotion Detection Model based on Sequential Neural Networks in Smart Exhibition Environment

Min Kyu Jung
School of Management, Kyung Hee University
(*minkyuli@khu.ac.kr*)

Il Young Choi
School of Management, Kyung Hee University
(*choice102@khu.ac.kr*)

Jae Kyeong Kim
School of Management, Kyung Hee University
(*jaek@khu.ac.kr*)

In the various kinds of intelligent services, many studies for detecting emotion are in progress. Particularly, studies on emotion recognition at the particular time have been conducted in order to provide personalized experiences to the audience in the field of exhibition though facial expressions change as time passes. So, the aim of this paper is to build a model to predict the audience's emotion from the changes of facial expressions while watching an exhibit. The proposed model is based on both sequential neural network and the Valence-Arousal model. To validate the usefulness of the proposed model, we performed an experiment to compare the proposed model with the standard neural-network-based model to compare their performance. The results confirmed that the proposed model considering time sequence had better prediction accuracy.

Key Words : Emotion detection model; Valence-Arousal model; Sequential neural networks; Facial features

Received : December 21, 2016 Revised : December 21, 2016 Accepted : February 2, 2017
Publication Type : Regular Paper Corresponding Author : Il Young Choi

1. Introduction

Recently, development of information technology is changing the exhibition industry from off-line environment to smart environment. In other words, audiences get personalized service with smart devices such as smartphone and tablet PC which interact with exhibition systems when visiting an exhibition hall. Moreover, exhibitors acquire the valuable audience information through exhibition system. Such smart exhibition environment is important to increase exhibition performance by providing new types of content suitable to audience preferences. Therefore, in order to improve exhibit performance in the exhibition industry and increase audience satisfaction, research on personalized interaction techniques to respond to the audience's responses is required in the field of exhibition (Jang, 1999). A prime example of such interaction techniques is emotion recognition. In order to provide customer-centered services in the existing exhibition industry, it is very important to understand comprehensively the involved emotions, symbols, audience behavior, and so on. The need for research on emotion recognition in the exhibition industry has recently increased gradually (Jung and Kim, 2012; Ko et al., 2008; Shim and Kim, 2009). In addition, various research works to provide intelligent services depending on user emotion (determined by facial expression) have been performed in the academic community. Research on emotion recognition has also facilitated the provision of personalized services to the audience.

Since human emotion entails subjective perception, it involves the user's primary intent as derived from contextual information. Therefore, if such significant information were utilized properly, more-advanced exhibit services would become available to the audience. Furthermore, personal emotional information will be a great resource for exhibitors' marketing strategies. Then, using this information, exhibitors will be able to roll out customized services to their customers (Jung and Kim, 2012; Yoo et al., 2011). As a result, emotion detection has become currently very active research field (Altun and Polat, 2009). Therefore, the aim of this paper is to propose an intelligent model to predict the audience's emotion upon seeing the exhibits on the basis of their facial expressions. This model could allow exhibition organizers to reflect changes in the audience's aesthetic and cognitive status according to the exhibit situation. Ultimately, a new type of interaction service or marketing strategy can be implemented according to audience emotion (Park et al., 2012).

Existing research that has tried to predict emotional state from user facial expression has been conducted continually in many other fields. However, existing research has the following common limitation: to determine users' emotions, many previous studies used ANN models; specifically, most used the standard ANN without time sequence (Hsieh, 2011; Ko et al., 2008; Shin, 2007). The ANN is good for recognition of static patterns but limited in the case of patterns with the time series, such as emotional transitions, as it incorporates no concept of time (Choi et al., 1992). Thus, in order to overcome these limitations of existing research, we built a prediction model for audience emotions using a sequential neural network (SNN) that incorporates time sequence. The SNN used in this study correlates emotional changes with time information by forwarding the output values of a past network the network for the present time. Further, in order to verify validity, this method was compared with emotion prediction based on the standard ANN model.

The rest of this paper is organized in the following manner. In the next section, theoretical background regarding the various factors that organize the proposed model in this research will be presented. In Section 3, the new interactive system in smart exhibition proposed in this paper will be introduced, and its procedures will be described. In Section 4, we will describe the experimental design used to analyze empirically the usefulness of the proposed model; that section also presents the results of empirical analysis. Finally, conclusions are presented and the limitations and implications of this study are discussed in the last section.

2. Literature review

1.1 Facial recognition

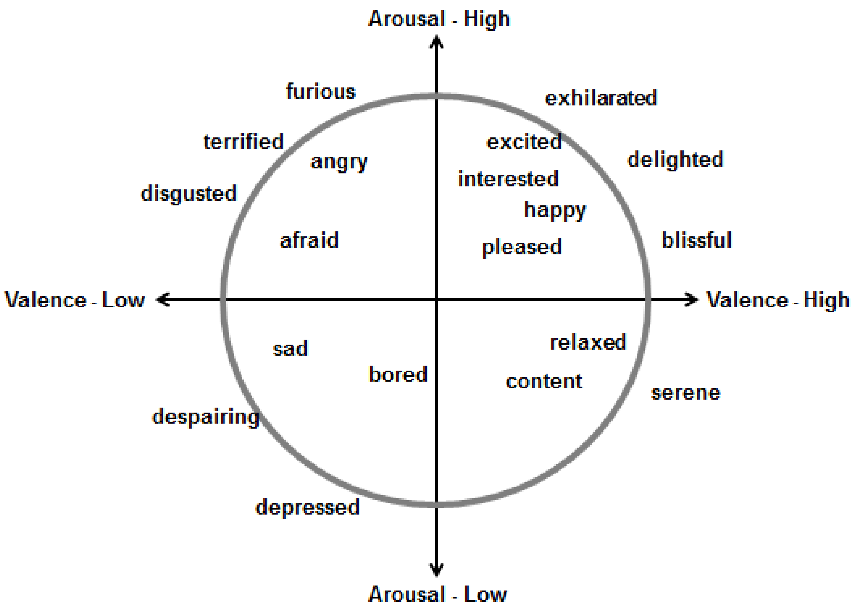
Human facial expression is an interesting target of exploration, because facial expression itself contains a variety of information (Wong and Cho, 2009). An attempt to organize research related to facial expressions, the Face Action Coding System (FACS), was proposed by Ekman (1972); it defines change in facial expression according to each facial feature and classifies these facial features as Action Units (AUs). Many previous studies have been based on FACS. Lee and Moon (2004) speculated that human expressions vary according to individual personality but also have certain similarities. In order to verify this, they took face pictures for the analysis of facial expressions. After that, they detected the common components of human faces from the face images. In addition, Park et al. (2000) analyzed facial expressions in terms of standard faces, including six kinds of emotion and mixing ratios between them.

1.2 Valence-Arousal (V-A) model

Attempts to classify and organize users' emotions have been performed through various studies incorporating models from Ekman's model to the V-A model. Existing research to determine emotions using user reactions almost utilized Ekman's 6-emotion classification system; that model classifies human emotions as six different types. Because the model has the advantage that it presents how change of expression is caused by each emotion, Ekman's emotion classification system has been applied by a number of studies to determine emotions through facial expressions. However, because these feelings are all negative emotions except for happiness, it may be difficult to take appropriate action even when the emotional determination is made accurately (Ekman, 1972; Oliveira et al., 2006).

Meanwhile, V-A model was proposed to measure human emotions with a two-dimensional approach (Mehrabian and Russell, 1974; Russell, 1980). The valence (V) dimension refers to how positive or negative the emotion is; it ranges from unpleasant feelings to pleasant feelings of happiness. The arousal (A) dimension refers to how excited or apathetic the emotion is, ranging from sleepiness or boredom to frantic excitement. Psychological evidence suggests that these two dimensions are inter-correlated (Alvarado, 1997; Lane and Nadel, 2000; Lewis et al., 2007; Ohno-Machado, 1996). More specifically, there exist repeating configurations and interdependencies within the values that describe each dimension. <Figure 1> shows emotion adjectives located in the two-dimensional V-A model.

Research to determine the status of users' feelings and emotions using the V-A model has been attempted in various fields. Nicolaou et al. (2011) determined V and A using an algorithm that used facial expressions, shoulder movement, sound, and so on as input variables. Yoo et al. (2009) sorted music using a V-A model whose

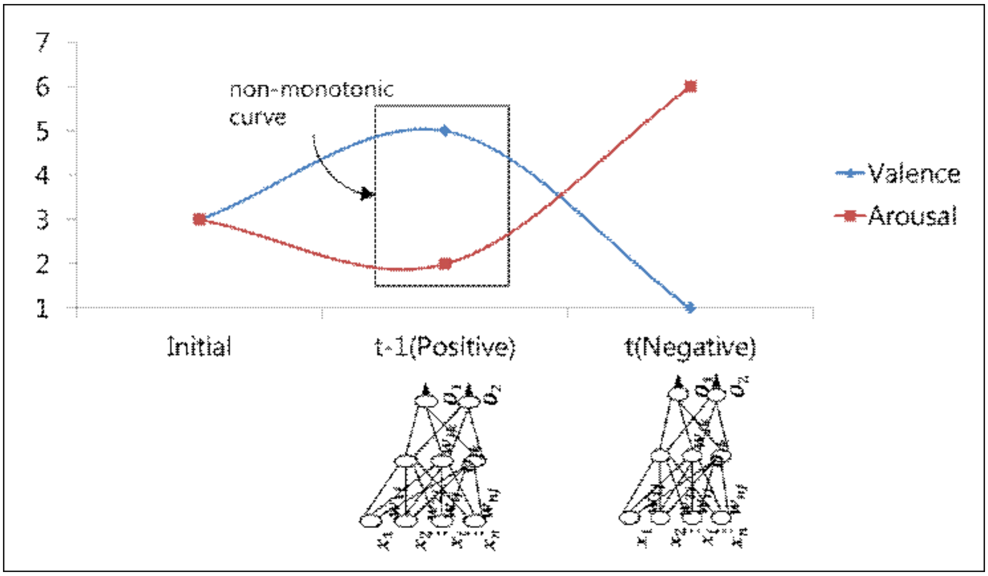


〈Figure 1〉 Valence-Arousal Model

input variable was sound. Then, based on that, they proposed a new interface for searching for music. Kim and Cho (2009) used the V-A model to express users' emotions as inferred by a Bayesian system. Further, they recommended the most appropriate ringtone in a mobile phone to users using inferred emotions, thereby facilitating user convenience. Accordingly, another purpose of this research was to organize audience emotions according to the V-A model, which is relatively more likely to be applied.

1.3 Standard neural networks

An artificial neural network (ANN) is a modeling technique based on the observed behaviors of biological neurons used to mimic the performance of the biological system. It is modeled as a network structure composed of nodes and links. ANNs are also used to explain patterns inherent in data through an iterative process of learning from past data (Park and Edington, 2001). Thus, standard ANNs are predominantly used to perform nonlinear input/output mappings and have become commonplace in many ANN applications. In other words, they are applied in a variety of fields, including pattern and character recognition, speech analysis, weather forecasting, robotics, and water/environmental engineering. Further, in the management field, standard ANNs are currently being used for a broad range of applications, such as those involving customer credit ratings, ferreting out bad deals, medical diagnostics, forecasting,



⟨Figure 2⟩ Standard Artificial Neural Network for a Single Point in Time

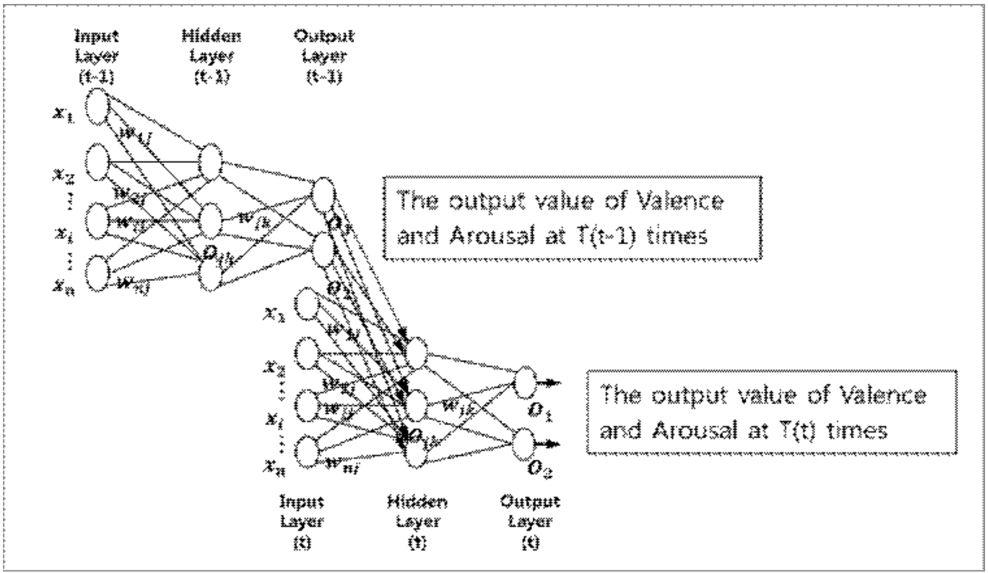
superior-customer selection, and targeted marketing. The standard ANN-based model is a fully connected feed-forward ANN with one hidden layer, one input layer, and one output layer (e.g., <Figure 2>).

For network training, the back-propagation algorithm was based on the minimization of error function; the average error was defined as the difference between actual and desired output (not depending on incremental time t). Moreover, it is used to implement validation for the prediction target using the learning results (Park and Edington, 2001).

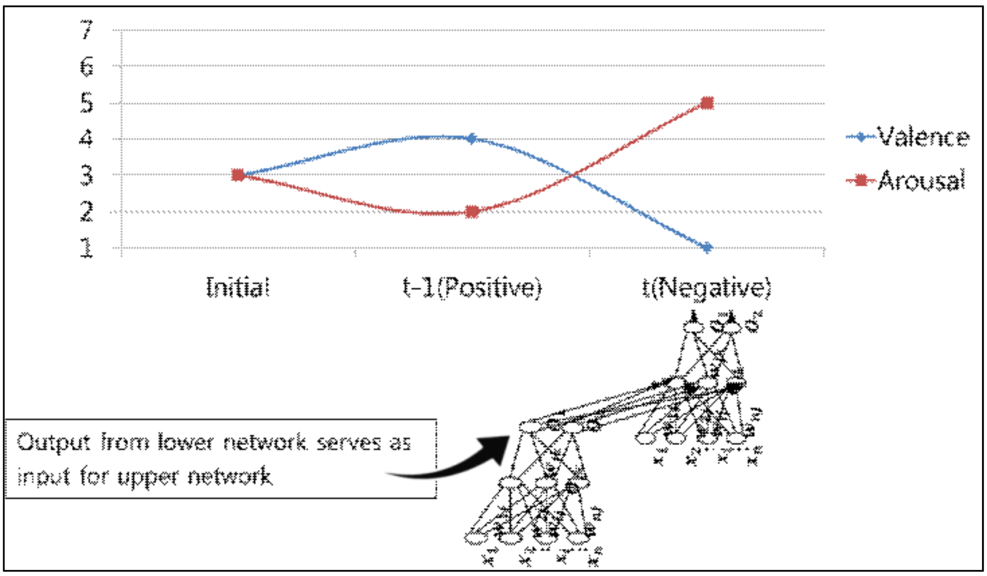
Studies that use standard ANNs to determine the user's emotion usually provide outcome predictions for a single point in time, as indicated in <Figure 2>. An attempt to build an emotion determination model using this standard ANN was made by Jung and Kim (2012) and Ryoo et al. (2013). However, Jung and Kim (2012) and Ryoo et al. (2013) built the emotion determination model for a single time point without considering temporal sequence. Therefore, this paper studied whether consideration of time sequence affects prediction accuracy of emotions.

1.4 Sequential neural networks

SNN systems are constructed incrementally (<Figure 3>), unlike standard ANNs. In each step of the sequence, predictions for one time point are produced by an ANN. These predictions are passed forward to other networks in the system. One or more networks may provide predictions that become inputs for other networks in the sequence. The result is a chain or a hierarchy of ANNs, as



〈Figure 3〉 Model Architecture of a Sequential Neural Network



〈Figure 4〉 Sequential Neural Network for Prediction

illustrated in <Figure 4> (Ohno-Machado, 1996; Ohno-Machado and Musen, 1997).

In SNNs, dependencies among time point predictions are represented explicitly, resulting in monotonic survival curves. In this example, predictions for time t − 1 are entered in the model that predicts emotion at time t. Generally, SNNs have been shown to provide predictions that are more accurate in terms of calibration and resolution than the ones produced by standard ANNs. SNN-based models have been used in various fields, particularly health care and the economy. Therefore, this paper applied this SNN-based technique to the field of smart exhibition to learn the emotion determination model. As a result, we confirmed that this SNN-based model has much better performance than the existing model based on a standard ANN.

3. Research model

The aim of this research is to propose a new type of emotion determination model to provide personalized services according to audience responses to smart exhibitions. This model determines the user's emotion using the change in facial expression among various audience responses. If this emotion determination model were developed, it would provide appropriate marketing strategies or personalized services for target users based on their emotions (as determined by the proposed model). Furthermore, effects such as increased sales and improved customer satisfaction can be expected.

Unlike existing research, in this paper, we improved prediction accuracy by considering time sequence. In other words, we used an SNN as a technique to learn the determination model. Through the experiment, this paper investigated whether time sequence affects human emotion.

The model proposed in this paper was constructed by the following procedure. The first step was data collection. In this step, data were selected to serve as the research model's inputs and outputs. The input values of this research model are comprised of the audience's facial feature values, and the output values are defined as V and A. Therefore, the first step is image capture of the audience members' faces to collect data on input values; then, the post-questionnaire task to collect data on output values will be performed. The second step is to preprocess the data collected in the first step. In this step, facial features are extracted from the collected data, and the model compensates for them. This step quantifies the position information on various facial features that configure human faces, like the eyes, nose, mouth, lips, eyebrows, and chin. The third step is to generate the independent and dependent variables of an ANN model, which served as the learning model of this research. In this step, the generated variables are preprocessed. In other words, the facial features obtained in the previous two steps are converted into concrete facial features during this step. The fourth step is to extract independent variables that affect the dependent variable using a statistical model. Statistical techniques like correlation analysis or independent-samples *t*-tests are mainly applied at this step. In the fifth step, the preprocessed data are divided into three data sets (the training, test, and validation sets). Further, during this step, two types of ANN-based models are learned: one is the standard ANN-based model, and the other is the SNN-based model. After the completion of this step, the construction of two ANN-based models to determine users' emotion is complete. Finally, the sixth step is to verify by using validation set whether the model constructed by the previous five steps reliably determines the emotions of the audience or not.

4. Empirical analysis

4.1 Data acquisition and preprocessing

In order to collect the data to build the emotion determination model, we participated in the 2011 DMC Culture Open held in Sangam-dong, Seoul, Korea. We collected the data from the event participants for three days. In total, 198 visitors participated in this event; demographic information about them is shown in the following <Table 1>.

〈Table 1〉 Subjects' Demographic Information

Characteristic		Subjects	
		The number of people	%
Gender	Male	101	51.0%
	Female	97	49.0%
Age	0–19 years	46	23.2%
	20–29 years	111	56.1%
	≥30 years	41	20.7%
Education	≤middle school	32	16.2%
	high school	33	16.7%
	university	88	44.4%
	graduate studies	45	22.7%
Job	student	104	52.5%
	office management	26	13.1%
	sales & service	4	2.0%
	engineering	4	2.0%
	manual labor	0	0%
	professional	27	13.6%
	self-employed	6	3.0%
	government employee	9	4.5%
	homemaker	14	7.1%
	unemployed	4	2.0%
Total		198	100.0 %

The number of data samples collected was 396, since we applied two videos to provoke different emotions in the participants. All of these images included the subject's face. Therefore, after 64 total facial points were extracted from these images, we calculated the values of facial features proposed by Pantic and Rothkrantz (2000) on the bases of these. A facial recognition solution from Olaworks (http://www.olaworks.com/), which is the leading domestic company for such solutions, was used to extract the facial points. In addition, the coordinates of the 64 facial points extracted from the images were corrected to the location of the fixed 30th coordinate set (i.e., the coordinates of the tip of the nose). Further, the data on V and A to be utilized as the dependent variables were collected by giving questionnaires directly to the audience members who participated in the experiment. Answers were rated on a 7-point Likert scale. The final data set of 264 total records used in the experiment was constructed through preprocessing; all were used in the experiment in this paper.

In order to build the model to predict V and A from the previously calculated dataset, characteristic independent variables having statistically significant correlations with each dependent variable were first selected. In this paper, we used two kinds of statistical techniques to select independent variables for use in the model. First, for independent variables on the ratio level of measurement, we acquired Pearson correlation coefficients through correlation analysis and verified whether they were statistically significant. Then, for independent variables measured as nominal (Boolean) variables, we extracted significant variables through independent-samples *t*-tests.

Let us examine the case of the standard ANN-based model. Under that model, in the cases of V and A, 10 and 8 out of 35 potential independent variables were determined to be statistically significant, respectively. In case of an SNN-based model, unlike the case with a standard ANN, these predictions for one time point produced by an ANN are passed forward to other networks in the system. In other words, the output values from the ANN model at time $t - 1$ become inputs for other networks in the sequence. Therefore, one output variable at time $t - 1$ is added to the input variables of the model for V and A at time t. As a result, 11 and 9 total independent variables were significant for V and A, respectively. <Table 2> contains a description of the finally selected independent variables.

4.2 Experimental design

In this paper, we built two models: both a standard ANN-based model and an SNN-based model. Then, we verified which one has better performance by comparing the two models. Before we built the two kinds of ANN-based models, we divided the 264-case data set into three subsets, as follows: the training, test, and validation sets. The distributions of the three data sets are shown as <Table 3>. The training, test, and validation sets were used to learn the determination model of

〈Table 2〉 Finally Selected Independent Variables

Name	Data type	Description of variables	Arousal	Valence	Unit
f1	Real number	angle ∠11,15,59		Selected(-)	Degree
f4	Real number	distance 23–64	Selected(-)		Pixel
f11	Real number	distance 32–37	Selected(+)		Pixel
f14	Real number	distance 32–34	Selected(-)		Pixel
f18	Real number	distance 32–56		Selected(-)	Pixel
AU1	Boolean	increased f1 & f2	Selected(+)		0:False 1:True
AU2	Boolean	increased f1 or f2	Selected(-)		
AU5	Boolean	increased f5 & f6		Selected(+)	
AU7	Boolean	decreased f7 or f8		Selected(-)	
AU10	Boolean	decreased f11		Selected(-)	
AU13	Boolean	decreased f12, decreased f13, decreased f14, decreased f15	Selected(-)		
AU15	Boolean	increased f12 or f13	Selected(+)		
AU18	Boolean	decreased f16	Selected(-)	Selected(-)	
AU20	Boolean	increased f16, non-increased f12, non-increased f13	Selected(+)	Selected(+)	
AU24	Boolean	decreased f17, decreased f16	Selected(-)		
AU41	Boolean	non-decreased f7, decreased f9, decreased f5 or decreased f10, decreased f6, non-decreased f8		Selected(-)	

audience emotion, explore the learning stop points in order to avoid overfitting, and verify the performance of the constructed model.

〈Table 3〉 Distribution of Data Set

Data set	ANN	SNN		Total
		T(t-1)	T(t)	
Training set	160	80	80	160
Test set	52	26	26	52
Validation set	52	26	26	52
Total	264	132	132	264

The experimental design for the building of the standard ANN-based model is as follows. To build the standard ANN-based model to predict V and A, this paper used the back-propagation algorithm as a learning method and configured both the learning and momentum rates at 10%. The sigmoid function was used as the activation function for transformation. Moreover, we designed a three-layer ANN with n input nodes, a number of hidden nodes, and one output node. Based on the test set, the learning was stopped at 50,000 iterations or after the minimum error was reached.

The number of nodes required to configure the hidden layer was selected through the following experiment. For V, we experimented with four cases having 5, 9, 14, and 18 nodes. Further, for A, we experimented with four cases having 5, 10, 15, and 20 nodes. After each experiment, a model was selected with a number of nodes corresponding with the one that showed the best performance. Neuroshell 2 4.0 was used as a tool to perform experiments using the ANN-based model.

In the case of an SNN-based model, an experiment essentially the same as the one using the standard ANN-based model was designed. However, unlike the existing standard ANN-based model, the SNN-based model consists of two ANN-based models: one is an ANN-based model at time t − 1, and the other is an ANN-based model at time t. When building models, first the ANN-based model at time t − 1 and then the ANN-based model at time t will be built. The model's output values at time t − 1 are used as the input values at time t. Therefore, the SNN model consists of a total of two ANN-based models.

4.3 Experiment results and analysis

In this paper, we used mean absolute error (MAE) as a measure to validate the prediction accuracy of the constructed model. MAE is the average of the absolute values of the differences between predicted and actual values; the smaller the MAE value, the higher the measurement accuracy. In this experiment, the performance of the measures of V and A were estimated using the validation set. The experimental results of building the standard ANN model for the prediction of V are as follows. We experimented with four cases having 5, 9, 14, and 18 nodes. Then, after each experiment, the number of nodes was adopted from the model that showed the best result. As a result, we obtained Table 4. As <Table 4> demonstrates, when the number of hidden nodes is 14, the prediction model for V has the highest accuracy. Thus, we selected the model based on the standard ANN model with 14 hidden nodes for prediction of V.

<Table 5> presents the experimental results for the standard ANN model to determine A. In this case, we experimented with four cases having 5,

〈Table 4〉 Experimental Results for Valence with ANN

Data set	h=5			h=9			h=14			h=18		
	T(t−1)	T(t)	Total	T(t−1)	T(t)	Total	T(t−1)	T(t)	Total	T(t−1)	T(t)	Total
Training set	0.8508	1.0517	0.9513	0.7962	1.1141	0.9551	0.7921	1.1622	0.9771	0.8026	1.1064	0.9545
Test set	1.1440	1.1084	1.1262	1.1351	1.1030	1.1191	1.0902	1.1842	1.1372	1.1279	1.1112	1.1196
Validation set	1.4414	1.1131	1.2773	1.418	1.1202	1.2691	1.4005	1.1168	1.2587	1.4147	1.1423	1.2785

〈Table 5〉 Experimental Results for Arousal with ANN

Data set	h=5			h=10			h=15			h=20		
	T(t−1)	T(t)	Total	T(t−1)	T(t)	Total	T(t−1)	T(t)	Total	T(t−1)	T(t)	Total
Training set	0.9369	1.1462	1.0415	0.9218	1.1059	1.0139	0.9533	1.1183	1.0358	0.9406	1.1113	1.026
Test set	0.8862	0.9358	0.911	0.84	0.9307	0.8853	0.8829	0.9517	0.9173	0.8301	0.946	0.8881
Validation set	0.8696	0.8274	0.8485	0.7741	0.8136	0.7939	0.7749	0.828	0.8015	0.7936	0.8182	0.8059

10, 15, and 20 nodes. Then, after every experiment, the number of nodes was adopted from the model that showed the best result. Experimental results showed that the determination model for A has the highest accuracy when the number of hidden nodes is 10. Thus, we selected the model based on the standard ANN model with 10 hidden nodes for prediction of A.

The following are the experimental results for the SNN-based model. This experiment was tested in four total cases, like the previous ones (<table 6> and <table 7>). First, the prediction accuracy was the best when the number of hidden nodes was 9 in the case of the SNN for V (<table 6>). Thus, we selected the prediction model for V based on the SNN with 9 hidden nodes.

〈Table 6〉 Experimental Results for Valence with SNN

Data set	h=5	h=9	h=14	h=18
	T(t)	T(t)	T(t)	T(t)
Training set	0.5999	0.5902	0.6135	0.4334
Test set	0.6079	0.6033	0.6094	0.6117
Validation set	0.5774	0.5750	0.5781	0.7134

Second, <Table 7> illustrates the case of the SNN for A; its prediction accuracy was best when the number of hidden nodes was 15. Thus, we selected the arousal prediction model based on the SNN with 15 hidden nodes.

〈Table 7〉 Experimental Results for Arousal with SNN

Data set	h=5	h=10	h=15	h=20
	T(t)	T(t)	T(t)	T(t)
Training set	1.1175	1.1264	1.1169	1.1074
Test set	0.9663	0.9195	0.9541	0.9336
Validation set	0.7888	0.7983	0.7851	0.8029

Finally, we compared the performance of the standard ANN and SNN models in determination of audience emotion. The results of the comparison are shown in <Table 8>. The table shows that the prediction accuracy of the model based on SNN is generally high. Specifically, In the case of V, we noticed that the prediction accuracy is much improved in the SNN model (<Table 8>). The reason for this result may involve the characteristics of the experimental data. In the

〈Table 8〉 Comparison between ANN and SNN

MAE of Valence	h=5		h=9		h=14		h=18	
	ANN	SNN	ANN	SNN	ANN	SNN	ANN	SNN
Training set	1.0517	0.5999	0.9457	0.5902	0.9463	0.6135	0.9460	0.4334
Test set	1.1084	0.6079	0.9736	0.6033	0.9738	0.6094	0.9699	0.6117
Validation set	1.1131	0.5774	0.8298	0.5750	0.8382	0.5781	0.8479	0.7134
MAE of Arousal	h=5		h=10		h=15		h=20	
	ANN	SNN	ANN	SNN	ANN	SNN	ANN	SNN
Training set	1.1462	1.1175	1.1059	1.1264	1.1183	1.1169	1.1113	1.1074
Test set	0.9358	0.9663	0.9307	0.9195	0.9517	0.9541	0.946	0.9336
Validation set	0.8274	0.7888	0.8136	0.7983	0.828	0.7851	0.8182	0.8029

dataset used in this experiment, the values of A contained in the data are generally higher than those of V. However, the values of V at time t − 1 are generally high because of the positive V of the stimulus presented at that time point, whereas the values at time t are relatively low because of the negative V of the stimulus presented then. This effect seems to be shown in the following results. In other words, the characteristics of the experimental data greatly affect the accuracy of the prediction model built on SNN. From these experimental results, we found that the SNN-based model outperformed the standard ANN-based model. In other words, we found that the prediction accuracy is improved when the time sequence is considered.

5. Conclusions

In this research, we built a model to determine a person’s emotional state from his/her facial expressions by taking advantage of the SNN technique with time series. Then, by applying the model to real data, we performed empirical analysis to measure the prediction accuracy of the proposed model. As a result, we confirmed that the SNN-based model had significantly better prediction accuracy than the existing model based on the standard ANN model. The achievement of this paper is expected to be used as the main theoretical basis for improvement of prediction performance when it is desired to take advantage of the emotion determination model in order to provide further personalized service during exhibitions.

The limitations and future research directions of this study are as follows. First, research on methods to improve the prediction accuracy of the emotion determination model using SNNs is needed. In this study, in order to determine the user's emotions at the current time t, the emotional state at the immediately previous time t − 1 was used. However, in future work, researchers could use the values at times t — 2, t — 3, and so on to determine whether they also affect user emotion. In addition, this paper used only facial expression as input for the emotion determination model. However, this information is not thought to be sufficient for the determination of audience emotions. Therefore, when different response data —such as sound and audience behavior—are used in addition to facial expressions, future research could improve the performance of the emotion determination model.

Second, the results of this research are difficult to generalize, because the model proposed in this paper is based on a single dataset collected during a particular event. Further, the levels of A contained in the data are generally high. Therefore, the prediction accuracy of A is less reliable than that of V. As a result, we would like to apply the proposed model to a more expanded dataset in future work. Further, it will be necessary to verify model performance closely using such data.

References

Altun, H., and G. Polat, "Boosting Selection of Speech Related Features to Improve Performance of Multi-class SVMs in Emotion Detection," *Expert Systems with Applications,* Vol.36, No.4(2009), 8197~8203.

Alvarado, N., "Arousal and Valence in the Direct Scaling of Emotional Response to Film Clips," *Motivation and Emotion*, Vol.21, No.4(1997), 323~348.

Eckman, P. "Universal and Cultural Differences in Facial Expression of Emotion." *Nebraska symposium on motivation*. Vol.19. University of Nebraska Press, Lincoln, 1972.

Haykin, S., *Neural Networks: A Comprehensive Foundation*. Prentice Hall, Upper Saddle River, NJ, 1999.

Hsieh, K.-L., "Employing a Recommendation Expert System based on Mental Accounting and Artificial Neural Networks into Mining Business Intelligence for Study Abroad's P/S Recommendations," *Expert Systems with Applications*, Vol.38, No.12(2011), 14376~14381.

Jang, N. S., *Data Mining*, Dae-chung Media, Seoul, Korea, 1999.

Jung, M. K., and J. K. Kim, "An Intelligent Determination Model of Audience Emotion for Implementing Personalized Exhibition," *Journal of Intelligence and Information Systems*, Vol.18, No.1(2012), 39~57.

Kim, D. S. *Neural Networks: Theory and Applications*. Hitech Information, Seoul, Korea, 1993.

Kim, Y. J., and S. B. Cho, "A Context-aware

Recommender System for Ringtones in a Mobile Phone using an Emotion Model", *Proceedings of Korea Computer Congress 2009 (KCC 2009)*, Vol.36, No.1(2009), 162~165.

Ko, K.-E., D.-J. Shin, and K.-B. Sim, "Development of Context Awareness and Service Reasoning Technique for Handicapped People," *Journal of Korean Institute of Intelligent Systems*, Vol.18, No.4(2008), 512~517.

Lane, R., and L. Nadel, *Cognitive Neuroscience of Emotion*, Oxford Univ. Press, London, UK, 2000..

Lee, C. H., and S. R. Moon, "Detection of Face-element for Facial Analysis," *Journal of The Institute of Electronics Engineers of Korea,-CI*, Vol.41, No.2(2004), 131~133.

Lewis, P. A., H. D. Critchley, P. Rotshtein, and R. J. Dolan, "Neural Correlates of Processing Valence and Arousal in Affective Words," *Cerebral Cortex*, Vol.17, No.3(2007), 742~748.

Mehrabian, A., and J. Russell, *An Approach to Environmental Psychology*, MIT Press, Cambridge, MA, 1974.

Nicolaou, M., H. Gunes, and M. Pantic, "Continuous Prediction of Spontaneous Affect from Multiple Cues and Modalities in Valence-arousal Space," *IEEE Transactions on Affective Computing*, Vol.2, No.2(2011), 92~105.

Ohno-Machado, L. "Sequential Use of Neural Networks for Survival Prediction in AIDS," *Proceedings of American Medical Informatics Association Annual Fall Symposium*, (1996), 170~174.

Ohno-Machado, L., and M. A. Musen, "Sequential versus Standard Neural Networks for Pattern Recognition: An Example using the Domain of Coronary Heart Disease," *Computers in Biology and Medicine*, Vol.27, No.4(1997), 267~281.

Oliveira, A. M., M. P. Teixeira, I. B. Fonseca, and M. Oliveira, "Joint Model-parameter Validation of Self-estimates of Valence and Arousal: Probing a Differential-weighting Model of Affective Intensity," *Proceedings of the 22nd Annual Meeting of the International Society for Psychophysics*, (2006), 245~250.

Pantic, M., and L. J. M. Rothkrantz, "Expert System for Automatic Analysis of Facial Expressions," *Image and Vision Computing*, Vol.18, No.11(2000), 881~905.

Park, J., and D. W. Edington, "A sequential Neural Network Model for Diabetes Prediction," *Artificial Intelligence in Medicine*, Vol.23, No.3(2001), 277~293.

Park, M. H., H. D. Ko, And H. R. Byun, "Analysis and Synthesis of Facial Expression using Base Faces", *Journal of KIISE: Software and Application*, Vol.27, No.8 (2000), 827~828.

Park, W. K., I. Y. Choi, H. C. Ahn, and J. K. Kim, "A Study on Intelligent Interactive System Considering Visitor's Response for Providing Personalized Exhibition Service," *Journal of the Korea Society of IT Services*, Vol.6, No.3(2012), 118~132.

Russell, J. A., "A Circumplex Model of Affect," *Journal of Personality and Social Psychology*, Vol.39, No.6(1980), 1161~1178.

Ryoo, E C., H. Ahn, and J. K. Kim, "The Audience Behavior-based Emotion Perdition

Model for Personalized Service," *Journal of Intelligence and Information Systems*, Vol.19, No.2(20123, 73~85.

Shim, H. G., and S. K. Kim, "The Study on the Forecasting of Operation Continuous Ability Applied the Artificial Neural Network in Battalion Defensive Operations," *Proceedings of Korea Institute of Industrial Engineers/The Korean Operations Research and Management Science Society*, (2008), 1~6.

Shin, D. I. Emotion Recognition Technology Trends, *Weekly Technology Trends*, Vol. 1283, (2007), 1~2.

Wong, J.-J. and S.-Y. Cho, "A Local Experts Organization Model with Application to Face Emotion Recognition," *Expert Systems with Application*, Vol.36, No.1(2009), 804~819.

Yoo, H., M.-Y. Kim, and O. Kwon, "Emotional Index Measurement Method for Context-aware Service," *Expert Systems with Applications*, Vol.38, No.1(2011), 785~793.

Yoo, M. J., H. J. Kim, and I. K. Lee, "Music Exploring Interface using Emotional Model," *Proceedings of 2009 HCI Korea Conference*, (2009), 707~710.

국문요약

스마트 전시환경에서 순차적 인공신경망에 기반한 감정인식 모델

정민규* · 최일영** · 김재경*

최근 지능형 서비스를 제공하기 위해 감정을 인식하기 위한 많은 연구가 진행되고 있다. 특히, 전시 분야에서 관중에게 개인화된 서비스를 제공하기 위해 얼굴표정을 이용한 감정인식 연구가 수행되고 있다. 그러나 얼굴표정은 시간에 따라 변함에도 불구하고 기존연구는 특정시점의 얼굴표정 데이터를 이용한 문제점이 있다. 따라서 본 연구에서는 전시물을 관람하는 동안 관중의 얼굴표정의 변화로부터 감정을 인식하기 위한 예측 모델을 제안하였다. 이를 위하여 본 연구에서는 시계열 데이터를 이용하여 감정예측에 적합한 순차적 인공신경망 모델을 구축하였다. 제안된 모델의 유용성을 평가하기 위하여 일반적인 표준인공신경망 모델과 제안된 모델의 성능을 비교하였다. 시험결과 시계열성을 고려한 제안된 모델의 예측이 더 뛰어남으로 보였다.

주제어 : 감정인식모델, V-A 모델, 순차적 인공신경망, 얼굴 특징점

논문접수일 : 2016년 12월 21일　논문수정일 : 2016년 12월 21일　게재확정일 : 2017년 2월 2일
원고유형 : 일반논문　교신저자 : 최일영

* 경희대학교 경영학부
** 교신저자 : 최일영
경희대학교 경영학부
1 Hoegi-dong, Dongdaemun-gu, Seoul 130-701, Republic of Korea
Tel: +82-2-961-9355, Fax: +82-2-961-0788 E-mail: choice102@khu.ac.kr

저 자 소 개

Min Kyu Jung

Min Kyu Jung(minkyuli@khu.ac.kr) obtained his MS at School of Management, Kyunghee University and his BS in Electronic Engineering from Kyung Hee University. His current research interests focus on Recommender Systems and business intelligence. He has published a paper which have appeared in Journal of Intelligence and Information Systems.

Il Young Choi

Il Young Choi(choice102@khu.ac.kr) obtained his MS and PhD at School of Management, Kyunghee University and his BS in Economics from Kyung Hee University. His current research interests focus on Recommender Systems, green business/IT, and business intelligence. He has published numerous papers which have appeared in International Journal of Information Management, Information Technology and Management, International Journal of Internet and Enterprise Management, Journal of the Korean Society for Management, Korean Management Science Review, Journal of Intelligence and Information Systems, and Information Systems Review.

Jae Kyeong Kim

Jae Kyeong Kim(jaek@khu.ac.kr) is a professor at School of Management, Kyunghee University. He obtained his MS and PhD in Management Information Systems (MIS) from KAIST (Korea Advanced Institute of Science and Technology), and his BS in Industrial Engineering from Seoul National University. His current research interests focus on business intelligence, network management, and green business/IT. He has published numerous papers which have appeared in Artificial Intelligence Review, Electronic Commerce Research and Applications, European Journal of Operational Research, Expert Systems with Applications, Group Decision and Negotiations, IEEE transactions on services computing, International Journal of Human-Computer Studies, International Journal of Information Management, Technological Forecasting and Social Change.