2025 IEEE/SICE International Symposium on System Integration (SII)
January 21-24, 2025, Munich, Germany

Automatic Emoticons Insertion System based on Acoustic Information of User Voice: 3rd Report on Hyper-parameter tuning of SVM and Visual Representation

Ryo Senuma, Sho Yokota, Akihiro Matsumoto,
Daisuke Chugo, Satoshi Muramatsu, Hiroshi Hashimoto, *Member, IEEE*

***Abstract*— This study proposes an Automatic Emoticons Insertion System that enhances text communication by estimating user emotions from speech and automatically inserting corresponding emoticons. The system uses Support Vector Machine (SVM) algorithms to classify emotions based on acoustic features extracted through the openSMILE toolkit. By optimizing SVM hyperparameters via grid search, the model achieved a significant improvement in classification accuracy, with the best performance at $C = 3.997$ and $\gamma = 0.03071$. Additionally, we introduce a visualization technique based on Plutchik's Wheel of Emotions to represent the emotional nuances conveyed by emoticons. This approach integrates emotion recognition into digital communication, improving emotional clarity and reducing misinterpretation.**

I. Introduction

With the widespread use of smartphones, text-based social media has become increasingly popular. As reported in [1], the LINE usage rate in Japan has seen continuous growth, surpassing 90% in FY2020, highlighting its crucial role in daily life. However, the rise in social media usage has also led to an increase in related issues. According to [2], a major concern is the misinterpretation of messages, which often stems from the lack of facial expressions and vocal tones on text-based platforms, making it difficult to understand the sender's emotions accurately. In Japan, emoticons and pictograms are commonly used to address this challenge. Fig. 1 displays examples of widely used Japanese emoticons. These symbols are effective in conveying emotions without audio-visual cues. However, some users have expressed frustration with having to open a separate dialog box to select and insert these emoticons into text, finding the process of inputting various data types to be time-consuming. This suggests that users prefer quicker communication methods over typing out text.

In this study, we propose a system that, in addition to converting speech input into text, uses acoustic information to estimate the user's emotions and automatically inputs emoticons at the end of the text that matches those emotions. It is assumed that the sender and receiver of the system are already acquainted with each other. To build this system, it is necessary to classify Japanese speech into emotional categories based on its acoustic information and select appropriate using grid search with the aim of improving the accuracy rate of emotion estimation based on acoustic information in Japanese. In addition, visualization of the estimation results considering the polysemous nature of emoticons and a method for emotion expansion is proposed.

This paper is organized as follows. In Section II, an overview of the system is provided. In Section III, we present the model voice data necessary for machine learning and the emotion labels to be classified, as well as the methods for extracting acoustic features and the machine learning algorithms employed. In Section IV, we discuss the relationship between each hyperparameter and accuracy and search for optimal hyperparameters through grid search tuning. In Section V, we report on the visualization of the estimation results and the extension of the model. Section VI summarizes the paper.

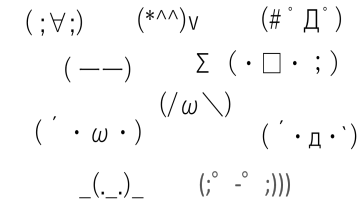

Figure 1. Example of Japanese Emoticons

II. Proposed System

Fig. 2 illustrates the workflow of the proposed system. To enable the automatic insertion of emoticons based on the user's emotions, an emotion estimation process is first required. The system uses the five basic emotions identified in the Atlas of Emotions [3] as the target emotions for estimation and representation. The Atlas of Emotions is a scientifically developed emotional map, created through the collaboration of the 14th Dalai Lama, the spiritual leader of Tibetan Buddhism and head of the Tibetan executive branch, and psychologist Paul Ekman. Initially, acoustic features of the speech are extracted using the open-source software openSMILE (Speech and Music Interpretation by Large-space Extraction) [4]. These features are then classified into one of the five emotions using a Support Vector Machine (SVM). Once these preliminary steps are completed, the speech is transcribed, the emotions are estimated, and an emoticon is selected corresponding to each emotion and inserted at the end of the text.

R. Senuma, S. Yokota, and A. Matsumoto are with Dept of Mechanical Eng., Toyo University, 350-8585 Kawagoe, Japan (e-mail: s36a02300135@toyo.jp, s-yokota@toyo.jp, akihiro@toyo.jp).
D. Chugo is with Kwansei Gakuin University, Japan (e-mail: chugo@kwansei.ac.jp).
S. Muramatsu is with Tokai University, Kanagawa, Japan (e-mail: muramatsu@tokai.ac.jp).
H. Hashimoto is with Advanced Institute of Industrial Technology, Tokyo, Japan (email: hashimoto@aiit.ac.jp)

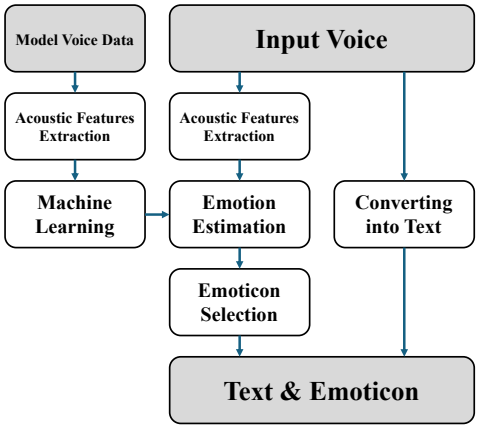

Figure 2. Proposed System Flow

III. Machine Learning Method

A. Model Voice Data

For emotion estimation, we utilize an emotional voice dataset [5] provided by Fast Label Corporation to learn the acoustic features associated with each emotion. This dataset includes professional voice actors expressing nine emotions (neutral, calm, joy, sadness, anger, fear, disgust, surprise, and impatience) through two different dialogue lines. In our previous study [6], the natural dialogue used in the model's voice data relied on emotion evaluations from three raters, which may have led to potential instability in the assessments. Additionally, in the acted voice recordings, the variation in dialogue lines for each utterance could have caused inconsistencies in the acoustic features within the same emotion. The emotional voice dataset provided by Fast Label Corporation addresses these issues by using consistent dialogue lines and recordings by professional voice actors, ensuring more uniform emotional expressions. In this dataset, the same dialogue is repeatedly uttered to express different emotions. This minimizes the variation caused by the different dialogue content and allows us to concentrate on analyzing the acoustic features corresponding to each emotion. Table I lists the five emotions based on the Atlas of Emotions that are estimated in this study, along with their definitions.

TABLE I. Five Emotions Definition

Emotion	Label	Definition
Joy	JOY	A feeling of strong happiness
Fear	FEA	A strong, bad feeling that you get when you think that something bad might happen
Sadness	SAD	The feeling of being sad
Disgust	DIS	A very strong feeling of dislike
Anger	ANG	The feeling that you want to shout at someone or hurt them because they have done something bad

B. Acoustic Feature Set

When creating the dataset, it is essential to extract acoustic features from the audio. For this purpose, we utilize openSMILE (open-source Speech and Music Interpretation by Large-space Extraction), an open-source software toolkit for speech analysis. openSMILE allows for the extraction of acoustic features, which are categorized into four main types of acoustic feature sets:

- ComParE_2016[7]

  The ComParE_2016 feature set comprises 6,373 static features, which arise from the calculation of diverse functionals over low-level descriptor (LLD) contours.

- emobase

  The emobase contains various acoustic features along with their first and second order derivatives. In addition, many statistical functions are applied to these features, resulting in a total of 988 features for every speech utterance.

- Geneva Minimalistic Acoustic Parameter Set (GeMAPS) [8]

  The GeMAPS has been conceived at an interdisciplinary meeting of voice and speech scientists in Geneva and further developed at Technische Universitat M¨ unchen (TUM). A total of 62 features are obtained for all utterances.

- extended GeMAPS (eGeMAPS) [8]

  An extended acoustic feature set for GeMAPS. A total of 88 features are obtained for all utterances.

In this study, we utilize the emobase as the acoustic feature set, based on the experimental results presented in the second report[9]. Furthermore, from the feature selection results using FFS in [9], a carefully selected set of 88 features is used. The list of features is shown in Table II.

C. Support Vector Machine

To perform emotion estimation, we use a machine learning algorithm known as Support Vector Machine (SVM). Compared to deep learning methods, SVM is a classical method, but it is computationally less costly. SVM is an algorithm designed to find the optimal hyperplane that classifies given data points into different classes, with this hyperplane being constructed to maximize the margin between classes. Additionally, for nonlinear data, the kernel trick can be employed to find complex classification boundaries. Based on the findings in the second report[9], the Radial Basis Function (RBF) kernel demonstrated high performance, and therefore, we use the RBF kernel in this study.

Python's scikit-learn library was used for training. Acoustic features were extracted from model audio data using openSMILE and labeled with five emotion categories. Subsequently, only the carefully selected 88 features were used, and standardization was performed. The RBF kernel SVM was employed for training and evaluation. In this study, 5-fold cross-validation was used. The dataset was divided into five subsets, with one used for testing and the remaining four subsets for training. This process was repeated five times, and

the accuracy of each fold was averaged to evaluate the model's generalization performance. This approach provides a more reliable performance assessment compared to evaluations that depend on a single training set.

TABLE II. FEATURE LIST SELECTED BY FFS

Acoustic Features	
pcm_loudness_sma_linregc1	lspFreq_sma[7]_linregerrA
pcm_loudness_sma_linregerrA	lspFreq_sma[7]_iqr1-3
mfcc_sma[1]_max	pcm_zcr_sma_amean
mfcc_sma[2]_linregc1	pcm_zcr_sma_linregc2
mfcc_sma[2]_linregc2	pcm_zcr_sma_linregerrQ
mfcc_sma[2]_stddev	pcm_zcr_sma_stddev
mfcc_sma[2]_iqr1-3	pcm_zcr_sma_kurtosis
mfcc_sma[3]_amean	voiceProb_sma_maxPos
mfcc_sma[3]_linregc1	voiceProb_sma_linregerrA
mfcc_sma[3]_linregc2	voiceProb_sma_skewness
mfcc_sma[4]_amean	voiceProb_sma_quartile3
mfcc_sma[4]_linregc2	F0env_sma_skewness
mfcc_sma[4]_linregerrA	pcm_loudness_sma_de_min
mfcc_sma[4]_quartile3	pcm_loudness_sma_de_linregc1
mfcc_sma[5]_stddev	pcm_loudness_sma_de_linregc2
mfcc_sma[5]_quartile3	pcm_loudness_sma_de_iqr1-3
mfcc_sma[6]_linregc2	mfcc_sma_de[1]_linregc1
mfcc_sma[6]_skewness	mfcc_sma_de[1]_quartile3
mfcc_sma[7]_amean	mfcc_sma_de[4]_linregc2
mfcc_sma[7]_linregc2	mfcc_sma_de[6]_linregerrA
mfcc_sma[7]_linregerrQ	mfcc_sma_de[9]_max
mfcc_sma[7]_quartile3	mfcc_sma_de[9]_linregc1
mfcc_sma[7]_iqr2-3	mfcc_sma_de[9]_linregc2
mfcc_sma[8]_max	mfcc_sma_de[9]_linregerrQ
mfcc_sma[9]_amean	mfcc_sma_de[10]_linregc1
mfcc_sma[9]_quartile3	mfcc_sma_de[10]_linregerrA
mfcc_sma[9]_iqr2-3	mfcc_sma_de[10]_linregerrQ
mfcc_sma[10]_iqr1-3	mfcc_sma_de[10]_stddev
mfcc_sma[11]_min	mfcc_sma_de[11]_linregc2
mfcc_sma[11]_linregc1	mfcc_sma_de[12]_iqr1-3
mfcc_sma[11]_stddev	lspFreq_sma_de[0]_max
mfcc_sma[11]_skewness	lspFreq_sma_de[2]_linregc1
lspFreq_sma[0]_linregc2	lspFreq_sma_de[3]_range
lspFreq_sma[0]_linregerrA	lspFreq_sma_de[3]_linregerrA
lspFreq_sma[0]_linregerrQ	lspFreq_sma_de[3]_stddev
lspFreq_sma[0]_kurtosis	lspFreq_sma_de[4]_stddev
lspFreq_sma[1]_range	lspFreq_sma_de[5]_linregc2
lspFreq_sma[2]_linregerrA	lspFreq_sma_de[7]_amean
lspFreq_sma[2]_kurtosis	lspFreq_sma_de[7]_linregc1
lspFreq_sma[4]_max	lspFreq_sma_de[7]_stddev
lspFreq_sma[4]_skewness	pcm_zcr_sma_de_linregerrA
lspFreq_sma[5]_amean	pcm_zcr_sma_de_skewness
lspFreq_sma[5]_linregc2	voiceProb_sma_de_linregerrA
lspFreq_sma[7]_linregc2	F0_sma_de_linregc2
Total of features : 88	

IV. HYPER PARAMETER TUNING

The RBF kernel SVM has two critical hyperparameters that are important for optimizing the model's performance. By tuning these two hyperparameters, it is possible to prevent overfitting and enhance the model's generalization performance. These two hyperparameters are:

- Regularization Parameter: C

  This parameter determines the size of the penalty for misclassified data. A larger value for C increases the penalty, making the model fit the training data more closely but increasing the risk of overfitting. Conversely, a smaller value for C reduces the penalty, allowing some misclassification and increasing the model's flexibility, but it also raises the risk of underfitting.

- RBF Kernel Parameter: γ

  This parameter controls the influence range of the RBF kernel. A larger γ value results in a more complex decision boundary that can better capture local structures but increases the risk of overfitting. Conversely, a smaller γ value smooths the decision boundary, making it more challenging to capture the overall patterns in the data.

A. Experimentation

In this study, we used grid search to find the optimal values for each hyperparameter. Grid search is a method that tests all predefined combinations of hyperparameters. To find the optimal values, we performed grid search using the following three settings:

- Setting 1

$$C = \{1,10,100\}$$
$$\gamma_i = 0.01 + \frac{(i-1) \times (1.0 - 0.01)}{49} \text{ for } i = 1,2,\dots,50$$

- Setting 2

$$C = 10$$
$$\gamma_i = 0.02 + \frac{(i-1) \times (0.04 - 0.02)}{99} \text{ for } i = 1,2,\dots,100$$

- Setting 3

$$C_i = 1.0 + \frac{(i-1) \times (500.0 - 1.0)}{999} \text{ for } i = 1,2,\dots,1000$$
$$r = 0.03071$$

In Setting 2, we narrowed the range based on the highest generalization performance value obtained in Setting 1 to further search for the optimal γ value. In Setting 3, based on the γ values explored in Setting 2, we searched for the optimal C value. Subsequently, we compared the generalization performance of the learning model with the optimized C and γ (to four significant figures) against that of the learning model using default settings.

B. Result and Discussion

- Setting 1

  The results for Setting 1 are shown in Fig. 3. The highest generalization performance of 71% was achieved when $C = 10$ and $\gamma = 0.030204082$. Additionally, since generalization performance decreases as γ increases regardless of the value of C, it can be inferred that C and γ are independent values.

- Setting 2

  The results for Setting 2 are shown in Fig. 4. The highest generalization performance of 71.09375% was achieved when $C = 10$ and $\gamma = 0.030707071$.

- Setting 3

  Among the results for Setting 3, the most notable change in generalization performance within the range $1 \leq C \leq 10$ is shown in Fig. 5. The highest generalization performance of 71.375% was achieved when $C = 3.996996997$ and $\gamma = 0.030707071$.

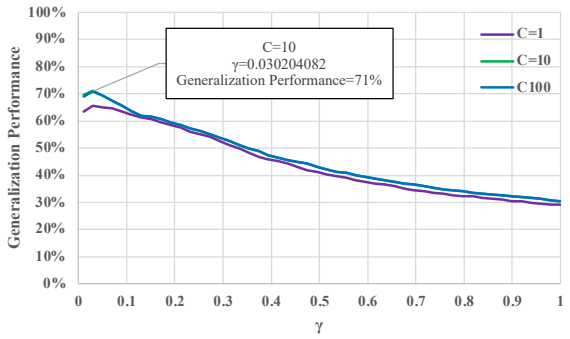

Figure 3. Grid Search Results for Parameter Setting 1

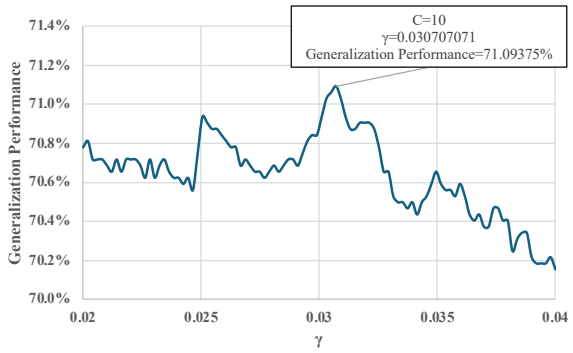

Figure 4. Grid Search Results for Parameter Setting 2

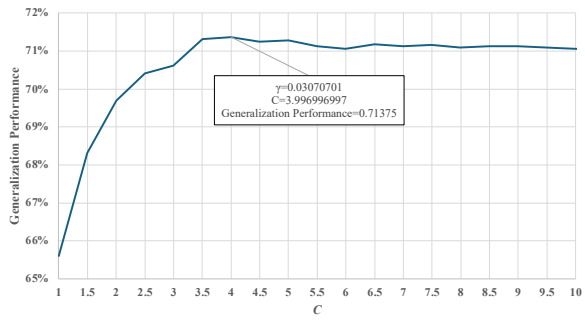

Figure 5. Grid Search Results for Parameter Setting 3

From the results above, the optimal value for the regularization parameter was $C = 3.997$, and the optimal value for the RBF kernel parameter was $\gamma = 0.03071$. The comparison of cross-validation results between the model trained with these optimal values and the model with default settings is shown in Fig. 6. The generalization performance of the default setting model was 64.97%, while the model with optimized parameters achieved a performance of 71.34%. This indicates that tuning the model's parameters significantly contributes to performance. Specifically, optimizing the regularization parameter improved the balance between overfitting and underfitting, allowing the model to learn a more generalizable boundary. Additionally, adjusting γ allowed the RBF kernel to better capture the appropriate nonlinearities in the input features, enhancing the model's flexibility and accuracy.

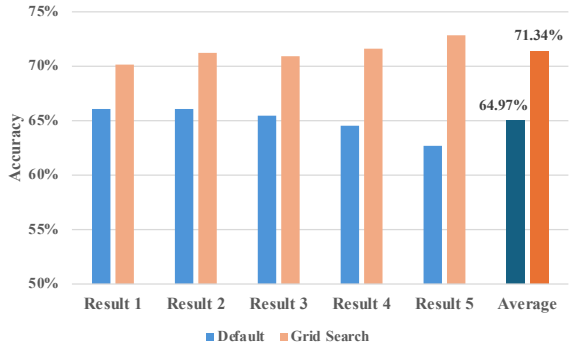

Figure 6. Comparison of Cross-Validation Results: Grid Search vs. Default Hyperparameters

V. Visual Representation of Estimation Results

The objective of this study is to automatically input appropriate emoticons based on estimated emotions. Emoticons are used to represent specific emotions, but a single emoticon can represent multiple emotions. For example, (> <) can be used to express sadness, surprise, or fear. This ambiguity provides flexibility in emotional expression but may impact the accuracy of emotion estimation. By using Plutchik's Wheel of Emotions[10], a model that visually represents emotional theories, we can geometrically understand that these emotions are closely related.

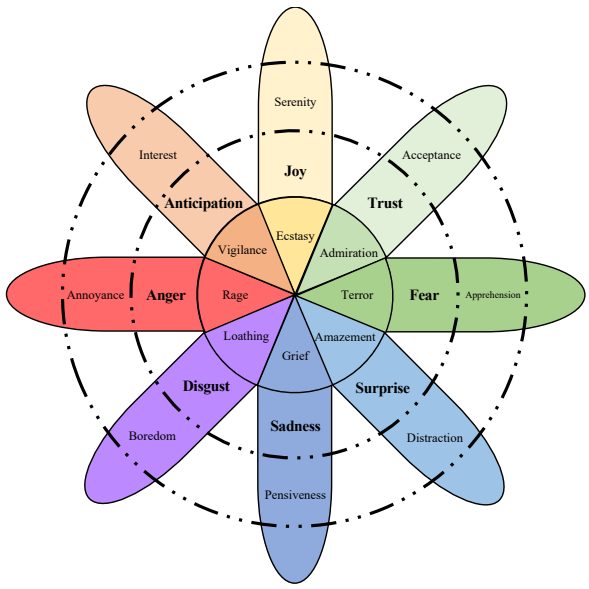

Figure 7. Plutchik's Wheel of Emotions

In the proposed emotion classification method, the input audio is classified into five emotions. To prevent misunderstandings on text-based social media, this study considers the polysemous nature of emoticons and expands these classification results to the eight emotions represented based on Plutchik's Wheel of Emotions. Fig. 7 shows Plutchik's Wheel of Emotions. This model categorizes basic human emotions into eight types (Joy, Trust, Fear, Surprise, Sadness, Disgust, Anger, Anticipation) and arranges them in a circular format to visually represent the polarity and proximity of each emotion. Since the five emotions that can be estimated by the proposed method are included in Plutchik's eight basic emotions, the estimation results can be visually represented on these eight emotions.

A. Circular Emotion Map

For the aforementioned expansion, the first step is to use the constructed RBF kernel SVM model to obtain the decision function values for each emotion class from the input data. The decision function values indicate how far the data point is from the classification boundary. Let $\mathbf{x}$ be the input data point (feature vector), $\mathbf{x}_e$ be the support vectors, a_e be the Lagrange multipliers, y_e be the class labels of the support vectors, $K(\mathbf{x}_e, \mathbf{x})$ be the kernel function, and b_e be the bias term. The decision function d_e is expressed as shown in (1). Here, the subscript e represents the five emotions listed in Table 1, $e = \{1: \text{FEA},\ 2: \text{JOY}, 3: \text{ANG}, 4: \text{DIS}, 5: \text{SAD}\}$.

$$d_e = f_e(\mathbf{x}) = \sum_{e=1}^{5} a_e y_e K(\mathbf{x}_e, \mathbf{x}) + b_e \tag{1}$$

Next, we create a circular emotion plane based on Plutchik's Wheel of Emotions. For this purpose, the circle of the emotional ring is created. The radius r of this circle is determined by adding 1 to the maximum value among the five values of decision functions d_e and then rounding down to the nearest integer, as shown in (2).

$$r = \lfloor \max(d_e) + 1 \rfloor \tag{2}$$

Additionally, to plot the eight basic emotions as defined in Plutchik's Wheel of Emotions evenly spaced around the circumference of a circle with radius r, we define the coordinates (x_i, y_i) of the basic eight emotions as shown in (3).

$$(x_i, y_i) = \left(r \cdot \cos\left(\frac{2\pi(i-1)}{8}\right), r \cdot \sin\left(\frac{2\pi(i-1)}{8}\right) \right) \tag{3}$$

Here, i is defined as shown in (4).

$$i = \begin{cases} 1, \text{if emotion} = \text{Fear} \\ 2, \text{if emotion} = \text{Trust} \\ 3, \text{if emotion} = \text{Joy} \\ 4, \text{if emotion} = \text{Anticipation} \\ 5, \text{if emotion} = \text{Anger} \\ 6, \text{if emotion} = \text{Disgust} \\ 7, \text{if emotion} = \text{Sadness} \\ 8, \text{if emotion} = \text{Surprise} \end{cases} \tag{4}$$

Furthermore, we define the range of each emotion using the plotted coordinates (x_i, y_i). The range corresponding to each emotion is defined as a 45-degree sector with (x_i, y_i) as the midpoint of the arc. The angle θ_i for each emotion is defined as shown in (5).

$$\theta_i = \text{atan2}(y_i, x_i) \tag{5}$$

Fig. 8 illustrates the constructed emotion circle and the positions of the eight basic emotions on it.

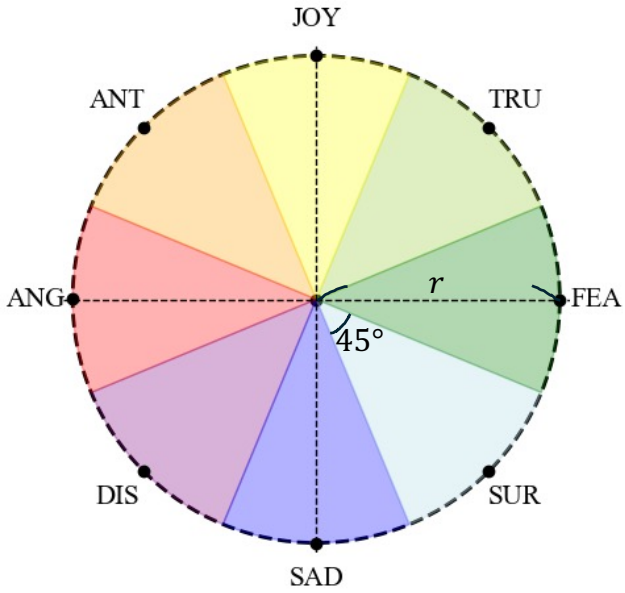

Figure 8. Circular Emotion Map based on Plutchik's Wheel

B. Plotting of Estimation Results

Using the decision function values of the five emotions estimated by the emotion estimation model, we map the input audio onto the circular emotion map defined by the eight basic emotions shown in Fig. 8.

First, the five decision function values derived from the estimation results are converted into coordinates and plotted on the circular emotion map that we created. If the value d_e of the decision function is considered as the distance from the center of the circle, the five coordinates (x_e, y_e) can be expressed as shown in (6). Here, if d_e is less than or equal to 0, it is set to $d_e = 0$.

$$(x_e, y_e) = \begin{cases} \left(\max(d_{fea}, 0), 0\right) \\ \left(0, \max(d_{joy}, 0)\right) \\ \left(-\max(d_{ang}, 0), 0\right) \\ \left(-\dfrac{\max(d_{dis}, 0)}{\sqrt{2}}, -\dfrac{\max(d_{dis}, 0)}{\sqrt{2}}\right) \\ \left(0, -\max(d_{sad}, 0)\right) \end{cases} \tag{6}$$

Next, create a polygon from the points plotted based on the estimation results. The coordinates of the weighted centroid (x_{center}, y_{center}) are calculated using the weights d_e for each emotion, as shown in (7).

$$(x_{center}, y_{center}) = \left(\frac{\sum_{e=1}^{5} d_e \cdot x_e}{\sum_{e=1}^{5} d_e}, \frac{\sum_{e=1}^{5} d_e \cdot y_e}{\sum_{e=1}^{5} d_e} \right) \tag{7}$$

A randomly selected data point (Emotion label: Disgust) not included in the training data of the emotion classification

model (the five emotions in Table 1) was used as input, and the calculated weighted centroid coordinates (x_{center}, y_{center}) were plotted on the circular emotion map. The coordinates of the calculated weighted centroid (x_{center}, y_{center}) were plotted on the circular emotion distribution map. The values of the five decision functions d_e at this time, according to (1), were $d_{fea} \cong 0.8326, d_{joy} \cong -0.2439, d_{ang} \cong 2.9827, d_{dis} \cong 4.2669, d_{sad} \cong 1.9392$. Therefore, according to equations (6) and (7), the coordinates of the weighted centroid $(x_{center}, y_{center}) = (-2.10, -1.66)$. Fig. 9 shows the circular emotion map with the weighted centroid $(-2.10, -1.66)$ plotted on it. From the plane, it is evident that even for input audio with an emotion label not included in the training data, the input is correctly plotted on the plane of the eight basic emotions, and the distribution of the input emotion can be visually confirmed. This demonstrates that using the five-emotion classification model, the system can accurately plot audio data not included in the training data onto the circular emotion map of eight basic emotions.

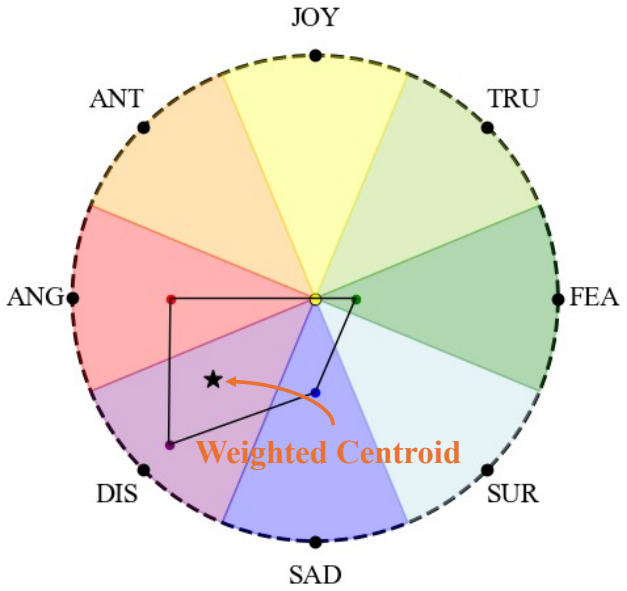

Figure 9. Weighted Centroid and Emotion Scores on Circular Emotion Map

VI. Conclusion

The objective of this study is to develop a system that estimates users' emotions from acoustic information and automatically inputs appropriate emoticons. This paper describes the hyperparameter tuning of the emotion estimation model using grid search to improve accuracy. As a result, it was confirmed that the optimal parameters are the regularization parameter $C = 3.997$ and the RBF kernel parameter $\gamma = 0.03071$. This optimization significantly improved the model's generalization performance compared to the default settings. Additionally, to visually extend the emotion estimation results, we proposed a method that creates a circular emotion map based on Plutchik's Wheel of Emotions and plots the coordinates of the estimation results. This method allows for a visual understanding of multiple emotion estimation results, enhancing the interpretability of emotion recognition.

In the future, we plan to explore weight adjustments that improve the accuracy of emotion estimation using weighted centroids and investigate the distribution of emoticons for each emotion. Our goal is to realize a real-time implementation of the proposed system.

References

[1] MIC Information and Communications Policy Institute, "Report on Survey on Information and Communication Media Usage Time and Information Behaviour in FY2020," pp. 1–66, 2021.

[2] MIC Information and Communications Policy Institute, "Survey Research on People's Awareness of New ICT Services and Technologies for Solving Social Issues," pp. 1–36, 2015.

[3] Dalai Lama, "The Ekmans' Atlas of Emotions," [Online]. Available: https://atlasofemotions.org/. [Accessed: Aug. 25, 2024].

[4] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: the Munich versatile and fast open-source audio feature extractor," in Proc. 18th ACM Int. Conf. on Multimedia, 2010, pp. 1459–1462.

[5] FastLabel Inc., "Emotional voice data set," [Online]. Available: https://fastlabel.ai/news/20230727-emotional-voice. [Accessed: Aug. 25, 2024].

[6] R. Senuma, S. Yokota, A. Matsumoto, D. Chugo, S. Muramatsu, and H. Hashimoto, "Automatic Emoticons Insertion System Based on Acoustic Information of User Voice: 1st Report on Data Model for Emotion Estimation Using Machine Learning," in Proc. 15th Int. Joint Conf. Computational Intelligence, ISBN 978-989-758-674-3, ISSN 2184-3236, pp. 548–554, 2023.

[7] B. Schuller, M. Neumayer, M. Baird, J. Burgoon, S. D. Kaiser, and K. Lang, "The Interspeech 2016 Computational Paralinguistics Challenge: Deception, Sincerity & Native Language," in Proc. 17th Annual Conf. Int. Speech Communication Association (Interspeech 2016), vol. 8, ISCA, 2016, pp. 1–5.

[8] F. Eyben, K. Scherer, J. Kroschel, and S. Schuller, "The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing," IEEE Trans. Affective Computing, vol. 7, no. 2, pp. 190–202, Apr. 2015.

[9] R. Senuma, S. Yokota, A. Matsumoto, D. Chugo, S. Muramatsu, and H. Hashimoto, "Automatic Emoticons Insertion System Based on Acoustic Information of User Voice: 2nd Report on Feature Selection for Emotion Estimation Using Machine Learning," in Proc. 2024 IEEE Int. Conf. Industrial Technology (ICIT), pp. 1–6, Mar. 2024.

[10] R. Plutchik, "A General Psychoevolutionary Theory of Emotion," in *Theories of Emotion*, R. Plutchik and H. Kellerman, Eds. Academic Press, 1980, pp. 3–33.