Sequential Model for Mapping Compound Emotions in Indonesian Sentences

Aripin^{1*}, Wisnu Agastya²

Department of Informatics, Computer Science Faculty, Universitas Dian Nuswantoro Semarang E-mail : *¹arifin@dsn.dinus.ac.id, wisnu.aga@gmail.com *Corresponding author

Hanny Haryanto³

Department of Informatics, Computer Science Faculty, Universitas Dian Nuswantoro Semarang E-mail : hanny.haryanto@dsn.dinus.ac.id³

Abstract - This research proposes mapping Indonesian sentences with single and multiple structures into emotion classes based on a multi-label classification process. The result of this research can apply in various fields, including the development of facial expressions in virtual character animation. Applications in other fields are facial expression analysis, human-computer interaction systems, and other virtual facial character system applications. In previous research, the classification process used for emotion mapping was usually based only on the frequency of occurrence of adjectives. The resulting emotion classes are less representative of sentence semantics. In this research, the proposed sequential model can take into account the semantics of the sentence so that the results of the classification process are more natural and representative of the semantics of the sentence. The method used for the emotion mapping process is multi-label text classification with continuous values between 0-1. This research produces the tolerant-method that utilizes the error value to deliver accuracy in the model evaluation process. The tolerant-method converts the predicted-label, which has an error value less than or equal to the error-tolerant value, to the actual-label for better accuracy. The model used in the classification process is a sequential model, including one-dimensional Convolution Neural Networks (CNN) and bidirectional Long Short-Term Memory (LSTM). The CNN model generates feature maps of each input in a partial way. Meanwhile, bidirectional LSTM captures information from input data in two directions. Experiments were performed using test data on Indonesian sentences. Based on the experimental results, bidirectional LSTM can produce an accuracy of 91% in the 8: 2 data portion and error-tolerant of 0.09.

Keywords - Sequential Model, Mapping Compound Emotions, Sentence Semantics, Indonesian Sentences

1. INTRODUCTION

Non-verbal communication which includes facial expressions, voice intonation and other expressive movements plays an important role in human interactive communication [1]. In the animation industry, interactive games, and other virtual character animation developments require facial expressions with a high level of naturalness. Facial expressions with a high level of naturalness are able to reinforce information conveyed in interactive communication. In humans, emotions play an important role in determining facial expressions [2]. However, machines do not have emotions like humans, so the determination of facial expressions on machines is done through a different approach. Emotional mapping from verbal

communication (written and spoken) can be used as a foundation in developing facial expressions in animated characters [3]. Emotion is a reaction to someone or an event. There are six basic emotions used as benchmarks in this study, including anger, disgust, fear, happiness, sadness, and surprise [4]. Emotions in Indonesian sentences can be recognized through adjectives [5] and the order of the words in a sentence.

Bahasa Indonesia is the official language used by Indonesians to communicate with each other. The structure of the Indonesian language includes subject, predicate, object, and description [5]. There are two types of sentences in Indonesian based on the number of clauses or ideas, namely single and compound sentences [6]. A single sentence is a sentence that has only one clause. Meanwhile, compound sentences are sentences with more than one clause separated by liaison words. Compound sentence structures have the potential to have multiple emotions which refer to the sentence structure which has more than one idea.

Mapping of emotions from Indonesian sentences is generally carried out based on a multiclass classification process with basic emotions. This process can extract emotional opinions from an Indonesian sentence. Text classification is the process of categorizing data into predetermined classes. The text classification process based on emotion class using traditional models including KNN, SVM and Naive Bayes is proven to be able to extract emotional data from an Indonesian sentence [4]. However, the classification process only refers to the factor in the number of occurrences of words so that it is not able to represent the semantic meaning. Therefore, a classification model that can represent the meaning of a sentence is needed to solve this problem.

Deep learning is a more specific development of machine learning that uses multi-layer perceptron in the case of supervised-learning, unsupervised and regression. In deep learning, there is a sequential model that can extract information from data vertically and horizontally. Sequential models include Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) which are used to process sequential data. CNN is generally used for spatial data such as in the case of image processing and computer vision. However, there are studies that use CNN for Natural Language Processing (NLP) cases [7]. Meanwhile, RNN is a model specially designed for processing sequential data such as NLP. In the development of RNN, there are Gated Recurrent Units (GRU) and Long Short-Term Memory (LSTM) which can solve the problem of vanishing gradients in the RNN standard. In general, vanishing gradients causes the model to lose information when processing long sequences of data, this problem often occurs in the RNN standard. Each sequential model in deep learning can capture information from data based on the layout or sequence of the data. So that the features captured by the model can be based on the occurrence of words and the word order of a sentence[8-10].

In connection with the problems in this study, there are several studies that have been conducted regarding the mapping of dominant emotions from Indonesian sentences [5]. The results showed the combined emotions of the Bayes naive multinomial model and the dominant boundary equation. The naive Bayes multinomial function is to get the probability value for each emotion class based on the number of occurrences of the adjective. While the dominant limit equation serves to produce a threshold value as a determinant of the dominant emotional class. But emotion classes are based solely on the number of occurrences of the adjective. Another drawback is that there is no evaluation of the dominant emotions generated[11-13].

Another study that discusses the classification process of English sentences using the Convolution Neural Networks (CNN) model [7]. The results showed that the CNN model was able to classify sentences with better accuracy measurement results than using the machine learning method.

Based on the description above, this study is proposed to map the dominant emotions of Indonesian sentences using a sequential model which includes the sequential use of the CNN and LSTM models. The sequential model can extract features that are not only based on the frequency of word occurrences, but more emphasis on the word order of a sentence so as to form semantic meaning. Therefore, this proposed study is aimed at improving the results of previous studies and similar studies. One of the advantages of this research is the result of the classification process in the form of compound emotions classes with better accuracy measurement results and the resulting emotion classes are more natural (natural). Thus the results of this study can be used for the synthesis of realistic and natural facial animations.

2. RESEARCH METHOD

Sequential models which include the CNN model and the LSTM bidirectional model are used to map emotions from Indonesian sentences. The emotion mapping process is carried out based on a multi-label text classification process with six basic emotions. The stages in mapping emotions from Indonesian sentences are described in Figure. 1.



Figure 1. Stages of Emotion Mapping

The emotion mapping stage includes the acquisition of a dataset, then continues with the text pre-processing stage to process unstructured data. After the text pre-processing stage is complete, a vectorization process is carried out to convert the text data into a vector that represents the information from each sentence in the dataset. Then the data is separated into training data for the training model stage and evaluation model testing set data.

a. Data Acquisition

This study uses a dataset in the form of Indonesian sentences with single and compound sentence structures. These sentences are taken from various sources, such as books, magazines, newspapers and digital media. The total number of sentences used as training data is 1306 sentences. While a sample of some of the data used as test data is shown in Table I. Furthermore, the sentences in Table I are given labels in the form of continuous values in the range [0-1]. These values are the result of the justification of a linguist by considering the emotional level of

each label from each sentence in Table I. The results of the justification of the label value of each emotion are shown in Table II.

	TABLE I		
_	INDONESIAN SENTENCES AS EVALUATION		
No.	Indonesian Sentences		
#1	sungguh bagus sekali ulahmu tadi pagi		
#2	betapa terkejutnya aku mendengar berita bodoh itu		
#3	senang sekali dia tiba-tiba datang ke acara ulang tahunku		
#4	aku takut melihat kecoa di dinding, apalagi kalau dia terbang		
#5	teganya kalian meninggalkan aku sendirian di tempat sepi		
#6	aku terkejut melihat ayah datang membawa kado		
#7	kenapa berat badanku harus naik		
#8	lebih baik aku pergi daripada melihatmu disini		
#9	sekilas aku melihat sosok bayang putih melintas di kegelapan		

CLASS LABE	L IN CON	ITINUOUS \	/ALUE
		**	<i>a</i> 1

No.	Angry	Disgusted	Afraid	Нарру	Sad	Shocked
#1	0.75	0	0	0	0	0.25
#2	0.50	0	0	0	0	0.50
#3	0	0	0	0.50	0	0.50
#4	0	0.25	0.75	0	0	0
#5	0	0	0.75	0	0.25	0
#6	0	0	0	0.75	0	0.25
#7	0.25	0	0	0	0.75	0
#8	0.25	0	0	0	0.75	0
#9	0	0	0.50	0	0	0.50

In Table II, each emotion has a continuous value in the range [0-1] which represents the intensity of each emotion. The use of continuous values on labels serves to represent more variations in the intensity of each emotion. For example, if the discrete labels include "very low", "low", "medium", "high", and "very high" only group each intensity into 5 categories. Meanwhile, the "low" category of each sentence has a different value. Therefore, the use of continuous labels can represent different intensities even though they are both in the "low" category. Examples of representing intensity in a continuous value are [0.0, 0.12, 0.25, 0.3, 0.65... 1.0].

b. Text Preprocessing

In the case of NLP, text preprocessing is an important step to clean unstructured data such as text. This stage affects the features generated from the dataset. The text pre-processing stage in this study includes case-folding, filtering, stemming and unstemming. Case-folding is the stage of changing all the letters in the dataset to lowercase, the filtering stage includes filtering stop words and punctuation. Stemming is the process of converting each word into a root word by removing affixes. Whereas the unstemming process is the opposite of stemming, this process

can preserve each word in its original form by not turning it into a root word. In Indonesian sentences, there are several words that can have different meanings from the basic word, such as words "bertaruh" with "taruh" and word "cobaan" with "coba".

c. Vectorization

Vectorization or vectorization is the conversion step of each sentence in the dataset into a vector that represents the sentence. In neural networks architecture, the input accepted is only numbers. So, each word must be converted to a number. This process is illustrated in Figure 2.

	Word	Index	
	acara	1	
	aceh	2	
	adik	3	
	salah	111	
	sambil	112	
		•	
	•	•	
	zaman	237	
Sentences	→ " saya membeli	baju untuk ł	nadiah ulangtahun ibu "
Vectorization	on → [95 75 32 131 81 137 87]		
Padding	→ [95 75 32 131 8	1 137 87 0 0	0 0 0 0]

Figure 2. Vectorization

This process represents each unique word in the dataset as an integer value stored in the vocab dictionary [7]. Then, each sentence is converted into a one-dimensional vector based on the vocab dictionary. So that each sentence produces a vector that represents the sentence. However, each sentence in the dataset has a different length, while the sequential model requires a fixed length input. Therefore, each input vector is added with the value 0 "zero" in the padding process, the zero value does not represent any word in the vocab dictionary, so it will not change the information of the sentence.

d. Split Data

After the vectorization stage, 1306 one-dimensional vectors are generated. Then to do the training process and model evaluation. The data is separated into two parts, namely the training data (training set) and the test data (testing set) randomly. This process is illustrated in Figure 3.



Figure 3. Split Data Illustration

At this stage, the data were separated in 8: 2 portions until a total of 1,044 training data and 262 test data were shown in Table III.

	TABEL III TRAINING AND TEST DATA		
#Data	Size	Number of Data	
Training data	80%	1044	
Test data	20%	262	

The training data is used for the model training process, while the test data is used for the model evaluation process in predicting unseen data or data that has never been seen by the model. The study comparing the two includes CNN and bidirectional LSTM models to map emotions from Indonesian sentences based on a multi-label classification process. However, the classification process with a continuous value label cannot be evaluated using direct classification accuracy. To be able to evaluate accuracy, a special approach is needed to process the actual and predicted labels. In this study, predicted label which is denoted as \hat{y} is processed by a special method developed in this study. This method uses the error value as a tolerance limit. The method is denoted in Equation (1) and Equation (2).

$$\mathbf{e}(\hat{\mathbf{y}}) = |\mathbf{y} - \hat{\mathbf{y}}| \tag{1}$$

tolerance(
$$\hat{y}$$
)=(e(\hat{y})≤ err_tolerance)=y: \hat{y} (2)

Where y is the actual label, \hat{y} is the predicted label, and e (\hat{y}) is the absolute error of the predicted label against the actual label. When e (\hat{y}) is less than or equal to err_tolerance, then \hat{y} is updated with the value of y. The err_tolerance notation is a value between 0 and 1. Then to calculate the accuracy described in Equation (3).

$$accuracy(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}(y_i = tolerance(\hat{y}_i))$$
(3)

Where y is the actual label, \hat{y} is the predicted label, and n is the number of labels. The lower the err_tolerance value, the tighter the calculation of model accuracy will be. The error tolerance denoted as err_tolerance determines how tolerant y is to \hat{y} . If the error value between y and \hat{y} is within the tolerance limit, then \hat{y} will be returned to y. So that when determining the accuracy of the value of \hat{y} can be equal to the value of y.

e. CNN Model

CNN is a sequential model that is generally used for computer vision cases. However, this model can also be used in the case of NLP [7]. The CNN model architecture in this study

includes word embedding layer, one-dimensional convolution layer, max-pooling layer, fully connected layer (dense layer), and output layer which are illustrated in Figure 4.



Figure 4. CNN Model

In Figure 4, each sentence that has been processed in the previous steps includes text pro-processing and vectorization then entered as an input vector in the embedding layer, this layer produces an embedding matrix of each input vector. Then from each column in the embedding matrix a convolutional layer is carried out which produces feature maps. All feature maps generated are forwarded in the max-pooling layer to eliminate each feature in the maximum distribution. After a feature map is generated which is eliminated in the max-pooling layer, then each feature map is forwarded to each neuron fully connected layer using the ReLU activation function. In the fully connected layer a weighting process is carried out, while the ReLU activation function functions to convert negative values to zero values. The results from the layer are then transmitted in an output layer consisting of six neurons with a sigmoid activation function. The use of the sigmoid function is intended so that each neuron can produce a value in the range [0..1]. Each neuron represents its own respective basic emotion label.

f. Bidirectional Long Short-Term Memory Model (LSTM)

LSTM is one of the RNN developments to overcome the vanishing gradient that occurs because the length of the text data is too large. LSTM can capture data information sequentially from the beginning to the end of the data. In addition to the CNN model, this study also uses a bidirectional LSTM model which has an architecture including word embedding layer, bidirectional LSTM layer, fully connected layer, and output layer which is illustrated in Figure 5.



As in the CNN model described above, the LSTM bidirectional model in Figure 5 accepts vector input. The input vector is a sentence that has gone through the text pre-processing and

vectorization stages. Then the input vector is included in the embedding layer to produce an embedding matrix that represents each column in the input vector. Each column in the embedding matrix is passed in a bidirectional LSTM layer. This layer has two types of LSTM layers, namely forward LSTM which captures information from the beginning of the data and backward LSTM which captures information from the end of the data. Then the results from the two LSTM layers are combined to be transmitted in each neuron fully connected layer. In the fully connected layer, the weighting of each input received by the layer is carried out. Every negative value in the fully connected layer is converted to zero value using the ReLU activation function. Then the resulting output vector is continued in the output layer. In the output layer there are six neurons representing each emotion label.

g. Word Embedding

Each input data is processed in a word embedding layer. Word embedding functions to convert each input vector into an embedding matrix. This process maps each word in a higher vector dimension. Each i-th word in each sentence is denoted as $x_i \supseteq R^k$, where k is the dimension of the embedding matrix. The embedding matrix is denoted in Equation (4).

$$x = [x_1, x_2, \dots, x_l]$$
(4)

Where $x \in R^{l \times k}$ and I are the length of sentence [7]. The embdedding matrix can be updated as you train the model.

h. Convolutional Layer

Convolutional layer functions to extract features from the embedding layer in onedimensional convolutional operations. One-dimensional convolution process uses a kernel filter to detect the features of each data by generating a feature map. The convolutional operation is formulated in Equation (5).

$$m_j = f(w \cdot x_{i:i+s-1} + b)$$
(5)

Where $w \in \mathbb{R}^{s \times k}$, s is the filter size, notation \cdot is dot matrix operation, b is the bias value, f is a non-linear function, and $x_{i:i+s-1}$ is formulated in Equation (6).

$$x_{i:i+s-1} = x_i + x_{i+1} + x_{i+s-1}$$
(6)

Where + notation is the concatenation of vector lines. Furthermore, to produce a feature map formulated in Equation (7).

$$\boldsymbol{m} = [m_1, m_2, \dots, m_{l-s+1}] \in \mathbb{R}^{l-s+1} \tag{7}$$

In a one-dimensional convolution layer, there are n-filters that generate an n-feature map. So that all processes in the convolution layer are denoted in Equation (8).

$$M = [m_1; m_2; ...; m_n] \in \mathbb{R}^{n \times (l-s+1)}$$
(8)

Where M is the output matrix of the convolutional layer, and the semicolons represent the line vector.

i. Max-Pooling Layer

Feature maps generated from the one-dimensional convolutional layer are continued in the max-pooling layer. In the max-pooling layer each feature map is processed in a maximum distribution that produces the highest value for each feature map. This process helps speed up the training process without sacrificing important information from each data. The convolutional and max-pooling processes are illustrated in Figure 6.



Figure 6. Convolution and max-pooling layer

j. Long Short-Term Memory (LSTM) Unit

LSTM is an extension of the RNN standard which reduces the problem of vanishing gradient [7]. The problem of vanishing gradient makes the standard RNN fail to capture information in long sequence data. The LSTM layer captures the long-term correlation of long sequence data by generating input gates, forget gates, output gates, and current memory. The LSTM unit is illustrated in Figure 7.



Figure 7. LSTM unit

The LSTM layer recursively processes the embedding vector of each cell block based on the hidden state h_{t-1} as another input of x_t , where t-1 is the previous time and t is the current time. In a cell block, the input gate is denoted as i_t, forget gate is denoted as f_t, output gate is denoted as o_t, and current memory is denoted as c_t. In Equation (9) describes the operation on the LSTM cell block.

$$i_t = \sigma \left(W^{(i)} x_t + U^{(i)} h_{t-1} + b^{(i)} \right) \tag{9}$$

$$f_t = \sigma \Big(W^{(f)} x_t + U^{(i)} h_{t-1} + b^{(f)} \Big)$$
(10)

$$o_t = \sigma \left(W^{(o)} x_t + U^{(o)} h_{t-1} + b^{(o)} \right) \tag{11}$$

$$u_t = \tanh\left(W^{(u)}x_t + U^{(u)}h_{t-1} + b^{(u)}\right) \tag{12}$$

$$c_t = i_t \cdot u_t + f_t \cdot c_{t-1} \tag{13}$$

$$h_t = o_t \cdot \tanh(c_t) \tag{14}$$

Where σ refers to the sigmoid function, $W^{(i)}$, and $U^{(i)}$ as the weight in the learning process and the sign \cdot (period) refers to the multiplication operation.

k. Bidirectional Long Short-Term Memory (LSTM) Layer

Bidirectional LSTM is two layers of LSTM with two different directions including forward-LSTM and backward-LSTM. Forward-LSTM captures information based on forward direction, while backward-LSTM captures information in backward direction. This process is illustrated in Figure 8.



Figure 8. Bidirectional LSTM Layer



Figure 9. Bidirectional LSTM Layer Many-to-One

In Figure 8, each input x is gradually processed in a forward and backward LSTM. The output of each LSTM forward and backward is combined in the sigmoid function to produce h_t , where t refers to the current time. The output matrix of the bidirectional LSTM layer is passed on to the next layer, which is illustrated in Figure 9. Units marked in red represent backward-

LSTM and units in blue represent forward-LSTM. Each h_t output denoted in the green unit is continued to the last output. Then, the output matrix is continued to the next layer.

The multi-label classification process is responsible for producing a combination of emotions from each emotion class. In the single-label classification process, text documents are classified into one class only. Meanwhile, the multi-label classification process classifies text documents into more than one class [11]. In multi-label text classification by binary class, the sigmoid activation function works very well in the output layer [12]. The formula for the sigmoid activation function is presented in Equation (15).

$$\sigma(z) = \frac{1}{1+e^{-z}} \tag{15}$$

Where z is the input value, and the output of the sigmoid function is always in the range between 0 and 1. The output layer in this study uses six neurons with the sigmoid activation function. The use of sigmoid in the output layer aims to carry out a multi-class classification process consisting of six output neurons representing each of the basic emotions (anger, disgust, fear, happiness, sadness, surprise). The output range of each neuron is [0–1] which represents the intensity of each emotion [13-16].

3. RESULTS AND DISCUSSION

a. Training Model

After the data is separated in the split data stage, training data and test data are generated. In the training process the training data is processed in a model to match the model with the information in the training data. The model training process uses the Binary Cross Entropy (BCE) as a loss function which is denoted in Equation (16).

$$BCE = \frac{1}{n} \sum_{i=1}^{n} y_i \cdot \log \hat{y}_i + (1 - y) \cdot \log(1 - \hat{y}_i)$$
(16)

Where \hat{y}_i is the predicted label and y_i is the actual label, n is the number of labels. Loss function is used to calculate the difference between the predicted label and the actual label [15]. The training process produces weights that match the information from the training data. After the model training process is carried out an accuracy evaluation of each model that has previously been normalized Equation (3).

b. Model Evaluation

In the model evaluation process, an error tolerance value is used with a value between 0.01 and 0.09 to determine the comparison of the accuracy of each model. The results of the model evaluation are described in Table IV.

TABLE IV MODEL EVALUATION MEASUREMENT					
Model	Error	Accuracy			
Tolerance					
CNN	0.01	0.7538			
Bi-LSTM	0.01	0.8121			
CNN	0.02	0.8155			
Bi-LSTM	0.02	0.8537			

CNN	0.03	0.8461
Bi-LSTM	0.03	0.8461
CNN	0.04	0.8635
Bi-LSTM	0.04	0.8839
CNN	0.05	0.8733
Bi-LSTM	0.05	0.8946
CNN	0.06	0.8835
Bi-LSTM	0.06	0.9009
CNN	0.07	0.8899
Bi-LSTM	0.07	0.9069
CNN	0.08	0.8971
Bi-LSTM	0.08	0.9124
CNN	0.09	0.9026
Bi-LSTM	0.09	0.9141

Table IV shows the model accuracy based on the error tolerance value, the experiment was carried out nine times with a tolerance value from 0.01 to 0.09 to determine the comparison of the performance of each model at various types of error tolerance. The smaller the error tolerance value, the more difficult it is for the model to obtain high accuracy.

Based on experimental results, at an error tolerance of 0.06 the LSTM bidirection model has achieved an accuracy of 90%. Meanwhile, the CNN model yields an accuracy of 88%. Then at an error tolerance of 0.09 the LSTM bidirectional model produces an accuracy of 91%. These results indicate that the LSTM bidirectional model is superior in mapping emotions from Indonesian sentences.

c. Results of Mapping Emotion Classes

After going through the evaluation stage, the LSTM bidirectional model was used to map the compound emotions of each sentence. In this stage, the results of emotion extraction from sample sentences using the LSTM trained bidirectional model are illustrated in Figure 10.

In Figure 10, two emotions dominate sentence # 1, namely surprise (0.24) and anger (0.75). In sentence # 2, there are two predominant emotional classes (anger: 0.52, and surprise: 0.48). In sentence # 3, there are two predominant emotional classes (happiness: 0.45, and surprise: 0.55). In sentence # 4, there are two predominant emotional classes (fear: 0.74, and disgust: 0.26). In sentence # 5, there are two predominant emotional classes (fear: 0.76, and sadness: 0.24). In sentence # 6, there are two predominant emotional classes (happiness: 0.72, and surprise: 0.28). In sentence # 2, there are two predominant emotional classes (anger: 0.22 and sadness: 0.78). In sentence # 8, there are two predominant emotional classes (sadness: 0.78 and fear: 0.22). In sentence # 9, there are two predominant emotional classes (surprise: 0.52 and disgust: 0.48).

Based on the experimental results described above, the six neurons in the output layer that use the sigmoid function can produce emotion classes (more than one emotion class) which can be used as the basis for the formation of compound expressions. Using a continuous value for each label results in a dynamic and natural value. The probability value of each emotion class can be used as a basis for determining the point value of the associated feature (unit of action). The action unit (abbreviated: AU) is a feature point that functions to control facial muscle movement in the process of facial animation.



Figure 10. Classification Result

4. CONCLUSION

Based on the experimental results in this study it can be concluded that a multi-label classification using a continuous value on a scale of [0-1] can produce a dynamic and natural intensity level of any combination of emotions. Then, to perform a continuous value evaluation process in text classification, an error tolerance method is generated to produce accuracy based on error tolerance. The use of the Sequential Model which includes CNN and bidirectional LSTM shows high performance with an accuracy value of 91% in the 8: 2 data portion with an error tolerance of 0.09. This can happen because in the LSTM bidirectional model there are two LSTM layers that capture information from two directions.

Emotion classes resulting from the classification process in this study can be used to develop animation of multiple facial expressions. The process of classifying Indonesian sentences that produce more than one emotion class can be used as a basis for forming compound expressions. Compound facial expression animation is a combination of several dominant emotional classes. Compound facial expression animations can be formed based on

the Facial Action Coding System (FACS) which consists of a combination of several Action Units (AUs). Action units are facial muscle movements that form facial expressions.

REFERENCES

- T. I. Kusumawati, "Komunikasi Verbal Dan Nonverbal", J. Pendidik. dan Konseling, vol. 6, no. 2, pp. 83–98, 2016.
- J. E. Prawitasllri, "Mengenal Emosi Melalui Komunikasi Nonverbal", Bul. Psikol., vol. 3, no.
 1, pp. 27–43, 2016, DOI: 10.22146/bpsi. 13384.
- [3] Vivi Triyanti, Yassierli, Hardianto Iridiastadi, "Basic Emotion Recognition using Automatic Facial Expression Analysis Software", Jurnal Optimasi Sistem Industri, Vol. 18, No. 1, 2019, DOI:https://doi.org/10.25077/josi.v18.n1.p55- 64.2019.
- P. Ekman and D. Cordaro, "What is Meant by Calling Emotions Basic," Emot. Rev., vol. 3, no. 4, pp. 364–370, Oct. 2011, doi: 10.1177/1754073911410740.
- [5] W. Agastya and Aripin, "Pemetaan Emosi Dominan pada Kalimat Majemuk Bahasa Indonesia Menggunakan Multinomial Naïve Bayes (Mapping Dominant Emotion in Indonesian Compound Sentences Using Multinomial Naïve Bayes)", JNTETI (Jurnal Nas. Tek. Elektro dan Teknol. Informasi), vol. 9, no. 2, pp. 171–179, 2020, DOI: https:// doi.org/10.22146/jnteti.v9i2.157.
- [6] S. Widodo, "Bahasa Indonesia Menuju Bahasa Internasional", Badan Pengembangan dan Pembinaan Bahasa, Kementerian Pendidikan dan Kebudayaan, 2013. http://badanbahasa. kemdikbud.go.id/lamanbahasa/ content/bahasa -indonesiamenuju-bahasa-internasional [accessed Mar. 01, 2020].
- [7] Y. Kim, "Convolutional Neural Networks for Sentence Classification", in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 1746–1751, DOI: 10.1109/CLEI.2017. 8226381. P. Zhou, Z. Qi, S. Zheng, J. Xu, H. Bao, and B. Xu, "Text classification improved by integrating bidirectional LSTM with twodimensional max pooling", in COLING 2016 - 26th International Conference on Computational Linguistics, Proceedings of COLING 2016: Technical Papers, 2016, vol. 2, no. 1, pp. 3485–3495.
- [8] Aditya Jain, Gandhar Kulkarni and Vraj Shah, "Natural Language Processing", International Journal of Computer Sciences and Engineering (IJCSE), Vol. 6, Issue 1, pp. 161-167, 2018.
- [9] Greg Van Houdt, Carlos Mosquera and Gonzalo Napoles, "A Review on The Long Short-Term Memory Model", Artificial Intelligence Review (2020), springer nature. com, published 13 May 2020.
- [10] D. Rahmawati and M. L. Khodra, "Automatic multi-label classification for Indonesian news articles", 2015 International Conference on Advanced Informatics: Concepts, Theory, and Applications, 2015, DOI: 10.1109/ICAICTA. 2015.7335382.
- [11] Ivana Clairine Irsan and Masayu Leylia Khodra, "Hierarchical multi-label news article classification with distributed semantic model-based features", International Journal of Advances in Intelligent Informatics, Vo. 5, No. 1, pp. 40-47, 2019, DOI: https://doi.org/10. 26555/ijain.v5i1.168.

- [12] Hui-Huang Zhao and Han Liu," Multiple Classifiers Fusion and CNN Feature Extraction for Handwritten Digits Recognition", Granular Computing, springernature.com, published 22 February 2019.
- [13] Yingying Wang, Yibin Li, Yong Song, and Xuewen Rong, "The Influence of The Activation Function in a Convolution Neural Network Model of Facial Expression Recognition", Applied Sciences 2020 (www.mdpu.com/journal/applsci), DOI:10.3390/app10051897, Published 10 March 2020.
- [14] Feiping Nie, Zhanxuan Hu, and Xuelong Li, "An Investigation for Loss Functions Widely Used in Machine Learning", Communication in Information and System, Volume 18, Number 1, pp. 37-52, 2018.
- [15] Ivo Duntsch and Gunther Gediga, "Confusion Matrices and Rough Set Data Analysis", Journal of Physics: Conf. Series 1229 (2019), pp. 1-6, 2019, DOI:10.1088/1742-6596/1229/ 1/01205.
- [16] Ahmad Hoirul Basori, Hani Moaiteq Abdullah AlJahdali, "Emotional Facial Expression Based on Action Units and Facial Muscle", International Journal of Electrical and Computer Engineering (IJECE), Vol. 6, No. 5, October 2016, pp. 2478-2487, DOI: 10.11591/ ijece.v6i5.12135.