Abstract - Neural networks have found profound success in the area of pattern recognition. By repeatedly showing a neural network inputs classified into groups, the network can be trained to discern the criteria used to classify, and it can do so in a generalized manner allowing successful classification of new inputs not used during training. With the explosion of research in emotion in recent year, the application of pattern recognition technology to emotion detection has become increasingly interesting. Since emotion has become an important interface for the communication between human and machine, it plays a basic role in rational decision-making, learning, perception, and various cognitive tasks. Human’s emotion can be detected based on the physiological measurements, facial expression and vocal recognition. Since human shows the same facial muscles when expressing a particular emotion, therefore the emotion can be quantified. In this study, six primary emotions such as anger, disgust, fear, happiness, sadness and surprise were classified using Neural Network. Real dataset of facial expression images were captured and processed to prepare for Neural Network training and testing. The dataset was tested on Multilayer Perceptron with Backpropagation learning algorithm and Regression analysis. The experimental results reveal that Neural Network has a misclassification rate of 2.5% while Regression analysis yields a misclassification rate of 33.33%. 

Keywords: Emotion, Classification, Neural Network

I. INTRODUCTION

Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid predictions. The first and simplest analytical step in data mining is to describe the data. However, data description alone cannot provide an action plan. Therefore a predictive model is required. Classification problems aim to identify the characteristics that indicate the group to which each case belongs. This pattern can be used both to understand the existing data and to predict how new instances will behave. Data mining creates classification models by examining already classified data (cases) and inductively finding a predictive pattern. These existing cases may come from a historical database or from an experiment in which a sample of the entire database is tested in the real world and the results used to create a classifier. Sometimes an expert classifies a sample of the database, and this classification is then used to create the model which will be applied to the entire database.

Data mining involves the use of sophisticated data analysis tools, including statistical models, mathematical algorithms, and machine learning methods (algorithms that improve their performance automatically through experience, such as neural networks or decision trees). In this study, a Neural Network technique is used as a classification algorithm to determine the type of child’s emotion of each facial expression image.

Neural networks (NN) have found profound success in the area of pattern recognition. By repeatedly showing a neural network inputs classified into groups, the network can be trained to discern the criteria used to classify, and it can do so in a generalized manner allowing successful classification of new inputs not used during training. With the explosion of research in emotion in recent year, the application of pattern recognition technology to emotion detection has become increasingly interesting.

Emotion also known as mood [1], and has always been used for showing human’s feeling. Human’s emotion state can be detected by some methods such as physiological measurement (heart rate, blood volume, blood pressure, skin resistance or conductance level, electroencephalogram, papillary response, electroculogram, gastrointestinal motility, electromyogram, skin temperature, brain
potentials, and respiration rate), facial expression and vocal recognition [2]. [3] also indicate that a human’s emotion state generally can be expressed by facial expression. The human’s emotion can be quantified since all human has the same facial muscles during expressing an emotion ([4], [5]).

User may interact with the computer by icons which appear in the human computer interface. Some software allow interaction between the system and the user through an intelligent agent such as Bonzi Buddy and Microsoft Clipper. However, these agents only interact with the user based on predefined rule. More natural and adaptive interaction between the system and user cannot be obtained while these agents are unable to adapt to user’s emotion in order to engage the user with the system. The engagement between user and system can be taken place in an interactive and adaptive environment. This is in line with Kate Hone [6] study, which indicates that emotion recognition technology has improved the computer interaction.

Since emotion has become an important interface for the communication between human and machine, it plays a basic role in rational decision-making, learning, perception, and various cognitive tasks. In this study, six primary emotions such as anger, disgust, fear, happiness, sadness and surprise were classified using Neural Network. Real dataset of facial expression images were captured and processed to prepare for Neural Network training and testing.

II. EMOTION AND FACIAL EXPRESSION

Emotion can be defined by three main theories, including James-Lange theory, Canon-Bard theory and Cognitive theory [7]. James-Lange theory stated that an emotion is caused by the human physiological changes after being stimulated by its environment. On the other hand, Canon-Bard theory indicated that the rising of the emotions and the human physiological responds occurs simultaneously. The cognitive theory of emotion extended the James-Lange theory by stating that the overall environment gives human clues that help human interprets the state of arousal.

Emotion can also be detected by some facial components such as eyebrows and lip [8]. This is in line with the study detecting the facial expression by the distinctive clues on the appearance of the mouth, eyes, and eyebrows [9].

Facial expressions are one of the most natural, powerful, and immediate means by which people communicate their emotions [8]. Many facial expression studies were based on the Facial Action Coding System that was developed [10]. Facial Action Coding System is used for measuring and describing the facial behaviours [4]. All visually distinguishable facial movement can be explained by referring FACS which has the code of facial expression clues. Many set of possible facial expressions can be created by the combination of these action units.

The human face expresses emotions faster than people verbalize or even realize their feelings. Many psychologists have studied human emotions and there are many possible facial expressions, the same expression may have radically different meanings in different cultures. However, it is widely accepted that there are six universal expressions that do not change too much from culture to culture. These universal emotion expressions are happiness, sadness, disgust, anger, surprise and fear ([10], [11]).

III. COMPUTATIONAL EMOTION AND NEURAL NETWORK

Affective Computing is a field of study that concentrates on how computers understand, express reaction appropriately with regard to emotions. Since its beginning, affective computing brings emotion and machines into different context. For example, works on Elliot’s Affective Reasoner implements the simulated environment containing multiple agents with emotional states based on the cognitive theory model [12]. The study shows that machine, which is capable to understand human emotions will be able to develop sense of profiling.

Another example of affective computing is an adaptive algorithm for learning changes in user interests [13]. This application employs personalized information filtering system that relies on users’ feedback. Using the feedback information, the profile is modified such that it will be incorporated for future filtering task. Affective learning companion provides another success story using affective
interaction in educational fields. Assisted with pedagogical approach, affective learning companion try to understand students’ acceptance level via their emotional level towards certain subjects [14].

[15] developed the emotional states taxonomy. The taxonomy provides some insight into how emotional states and its transitions between them govern the agent’s personality. From the taxonomy, emotions space was divided into two parts, positive and negative affects. Each branch provides information about the valance, duration and event for emotional representation template. Another approach proposed by [16] focuses on computational emotion model using probabilistic approach. In this case, assumption about the world (environment) is important for personality development. The main advantage of using this approach is that a personality model can be learned, given limited amount of observations of the other agent’s behaviour [17].

The related works in emotions and Cognitive Science indicate that the need for developing life-like and socially sophisticated system increases. Research has also shown that engaging anthropomorphic agent for interacting with users has become an important part of the system functionality. In order to develop human like and socially sophisticated system, an understanding towards emotion modelling has become vital. In developing computational emotion modelling, several techniques have been discussed, however for an initial attempt, the emotion model developed by [10]; and [11] were used as a basis for this study.

NN has been used by facial expression researchers in order to learn the facial expression patterns. [18], [3], [19], and [9] used multilayer perceptron with back propagation learning algorithm to recognize facial expressions. The classification accuracies in these studies were approximately not less than 90%.

Other researchers like [20] and [9] used Radial Basis Function. In these studies, the classification accuracies obtained by RBF were ranging from 90% to 92.1%. Hopfield NN has also been applied to facial expression recognition problem. For example, [21] achieved 92.2% classification accuracy when used this network to identify four facial expressions.

IV. DATA PREPARATION

Data preparation is an important phase since the prepared dataset becomes input to the neural network training and testing. Once the image has been captured and extracted, the image processing techniques are applied on this image. The overall stage of image processing is illustrated in Fig. 1.

![Fig. 1. The Flow of Process of Image Processing](image_url)

The expression images were captured with the Sony Cybershot DSC U50 digital camera and the images are shown in Fig. 2.
Image filtering is then performed to process an image as a preparation for training and testing. The gray scale transformation is performed using the intensity component that is written in (1).

\[ I = \frac{1}{3} (R + G + B) \]  

(1)

where \( R \) represents Red  
\( G \) represents Green  
\( B \) represents Blue

The Contrast-Limited Adaptive Histogram Equalization was used to adjust and further enhanced the contrast, thus convert the filtered image to the equalized image.

For Thresholding stage, Otsu’s method was employed to capture the global image threshold by selecting the threshold that minimizes the intra-class variance of the black and white pixel [24]. The generated global threshold was used to convert an intensity image to a binary image that is represented as numerical value either 0 or 1 [25]. The data preparation stages and its effect on the images are shown in Fig. 3.

The second last phase of data preparation is the Edge Detection stage. The information with regard to three facial components: left eyebrow, right eyebrow, and lip that represent emotions were obtained from the simplified image. Since edge is always associated with the boundaries of the objects of an image, Sobel technique is used for detecting its edge. In this study, the Sobel automatically generates a suitable value from the image. If the edge is lower than this value, the edge value is ignored.

\[ |G| = \sqrt{G_x^2 + G_y^2} \]  

(4)

In order to transform the images into numerical values, the left and right eyebrows as well as the lip were considered. Based on initial explanatory study, each eyebrow is segmented to the three rows and seven columns. As a result, 21 boxes with 10 X 10 cells in each box are created. The lip images consists of five rows and eight columns. This gives rise to 40 boxes with 10 X 10 cells in each box. Each segmented image is transformed to numerical values by calculating the ratio of existing pixel cells in each box. Table 1 shows sample of segmented images that has been transformed to numerical values.
TABLE I
SAMPLES OF IMAGE DATA TRANSFORMED TO NUMERICAL DATA AND NUMBER OF CREATED DATA

<table>
<thead>
<tr>
<th>Form of Facial Component</th>
<th>Numerical Data</th>
<th>Number of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>21</td>
</tr>
<tr>
<td>Final Left Eyebrow Image</td>
<td>0.00 0.25 0.02 0.79</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.89 0.63 0.39 0.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.56 0.02 0.37 1.00</td>
<td></td>
</tr>
<tr>
<td>Final Right Eyebrow Image</td>
<td>0.76 0.16 0.00 0.53</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.30 0.83 0.00 0.68</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.40 1.00 0.00 1.00</td>
<td></td>
</tr>
<tr>
<td>Final Lip Image</td>
<td>0.53 0.30 0.50 0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.40 0.16 0.00 0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.21 0.00 0.00 0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.16 0.00 0.00 0.00</td>
<td></td>
</tr>
</tbody>
</table>

A total of 120 patterns has been collected and preprocessed to form dataset for NN training and testing. This dataset consists of six classes of emotion. Each emotion is represented by 82 variables, 21 of which represent the left eyebrow, the next 21 represent right eyebrow, and the last 40 variables represent lip. Each class of emotion is represented by the symbol, such as AG (Anger), SP (Surprise), HP (Happiness), SD (Sadness), FR (Fear), DG (Disgust).

V. RESULTS

Initial study has been conducted to determine the suitable number of epoch prior to conducting experiments for determining the backpropagation training parameters. For experimental purposes, the learning rate and the momentum rate were both set to 0.1. Based on the results exhibited in Fig. 4(a), epochs 300, 400 and 500 obtained the highest test results with 75% classification accuracy (25% misclassification rate). These three number of epochs were further investigated by varying weight seeds in order to determine the most suitable number of epoch for the problem at hand. The results displayed in Fig. 4(b) show that a set of network trained up to 500 epoch achieved the highest average test result with 60% accuracy.

In order to determine the most suitable number of hidden units, the dataset was trained with various hidden units ranging from 2 to 20. The results illustrated in Fig. 5 (a) indicate that hidden unit 8 and 10 obtained 100% accuracy for both training and testing. A set of networks with hidden unit 8 and 10 were further trained to determine which hidden is more appropriate to be used in the next experiment. The results depicted in Fig. 4(b) show that a network with 10 hidden units obtained higher classification accuracy than a network with 8 hidden units (97.50% versus 95.00%).
Similar experiments were conducted to determine the learning rate and the momentum rate for Backpropagation learning algorithm. The experimental results show that learning rate 0.1 obtained 94.59% classification accuracy whilst momentum rate of 0.1 achieved 97.50%. Once again the number of epoch was investigated based on the selected training parameters of Backpropagation. Based on the experimental results, the NN architecture is 82-10-6 with misclassification rate of 2.5 percent. After performing several other experiments, the final epoch for obtaining such a network is 200 epoch. Further analysis indicate that emotion of type Disgust is harder for NN to recognize.

The same dataset was tested with Regression analysis that yields 33.33% of misclassification accuracy.

VI. CONCLUSION

Future studies can be focused on extending the dataset used in the study. The sample images of different group of subjects with different ages and races may also be considered. The larger sample of images would in effect yields a more representative classification emotion model. Nevertheless, the emotion classification model developed in this study can support the development of Intelligent Tutoring System in particular, and E-learning system in general.

REFERENCES


