Emotion Learning: A Computational Model for the Limbic System

(https://github.com/zytyz)

Abstract

A large majority of current methods in artificial intelligence have been developed and outperform humans in certain tasks. However, most of these models do not take the human brain structure into account. In order to create more "human-like" machine instead of "rational" machine, emotion must be considered as a factor. This report focus on the emotions in human brain and computational models for emotions. One of the existed computational models for emotion recognition (BELPIC) is reviewed and implemented in python as well.

1. Introduction

Why should we understand emotions?

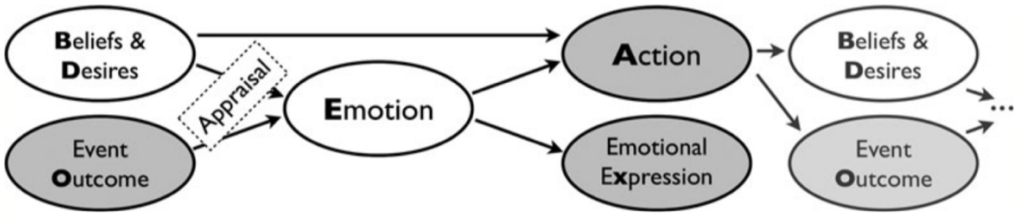
Artificial Intelligence and machine learning have progressed rapidly in recent years. A large majority of current researches focus on how to create, or teach a machine that outperforms human in different tasks. For example, Convolution Neural Networks with some certain training techniques are able to outperform human in object recognition. Some recurrent neural networks outperform human in text reading and semantic analysis. However, although these models seem to be "artificial intelligence", the underlying architectures are actually far from the human brain. One important mechanism they lack is emotion, which is a crucial to how humans act and make decisions since humans are not always rational. If we consider the importance of "emotional intelligence" in machine learning, we would be able create more human-like machine. Hence, in this report, I would like to focus on the mechanisms of emotions in the human brain and try to implement one of the computational models for emotion recognition – the "Brain Emotional Learning- Based Visual Pattern Recognizer".

2. Computational Models of Emotions

2.1 Framework for existing Models for Emotions

Many existed computational models try to understand how people reason about emotions, which can also be termed as "affective cognition" [2]. People have lots of intuitive theories of how other people might feel or how emotions affect their behavior. For example, when seeing another person crying, one might conclude that it is because the person is sad. People ("observers") use their intuitive theory of emotion to reason about the emotional states of others ("agents"). This intuitive theory depends on the observer's past history and experience. There are two types of causal relationship in an observer's intuitive theory of emotion: 1) What causes an agent to feel an emotion? 2) How does an agent react to an emotion?

Two factors cause emotions in an agent: an outcome of an event, and the agent's own beliefs and desire. For instance, a fan of a basketball team might experience happiness to see the team win. However, the same outcome would probably let fans of the opponent team feel anger or disappointment. An agent might also react to an emotion, which could be an emotional expression, such as facial expressions or body language, or actions that the agent might take. This action will then cause another outcome of an event, which would further cause another emotion and so on. All these causal relationships mentioned above can be drawn in a diagram (Fig. 1).



(Fig. 1) A model of an intuitive theory of emotion

Consider Fig. 1 as a Bayesian network. Each arrow could be treated as a conditional probability. For example, the arrow from *Emotion* to *Emotional Expression* stands for *P(Emotion |Emotional Expression)*, which is the probability of one's emotion when his/her emotion expression is known. This task is also known as **emotion recognition**, which is one of the popular research topic of emotions and lots of computational models have been proposed. Other arrows in the network are also research topics about emotions with related computational models. Together, the network is a framework for most existing computational models of emotions. The following of this report will be focused on emotion recognition particularly.

2.2 Emotion Recognition: Machine Learning Method

One of the most popular methods for emotion recognition is to treat it as an image classification problem. Studies have shown that there are 6 basic emotions for human beings: angry, disgust, fear, happy, sad, and surprise [3]. Hence, a CNN model for facial expression classification can complete the task. The input to the model are images of human faces with different facial expression, while the output is the emotion this image represents. However, this method does not take the human brain structure into account thus is unable to provide insights to how humans decode facial expressions. In order to gain more knowledge about how the brain recognize emotions, one should consider the neurobiology basics about emotions and include the mechanism in the proposed model.

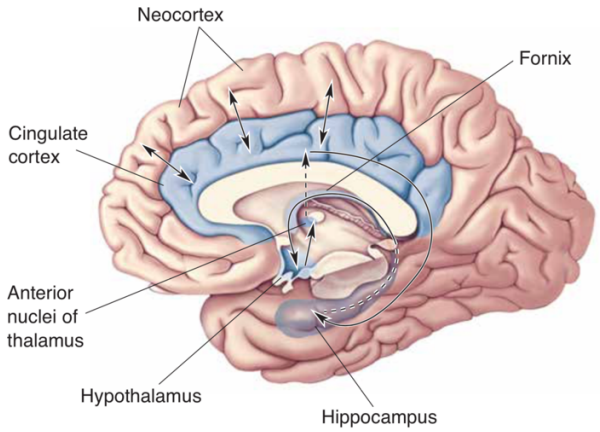
3. A Neural Basis Computation Model for Emotion Recognition

The "Brain Emotional Learning- Based Visual Pattern Recognizer" (BELPIC) is a

proposed computational model for emotion recognition based on the limbic system [4], which is the region of brain that plays a main role in emotion processing and response. The following section is arranged in three parts. The neurobiology for emotion recognition would first be introduced. A brief review on the BELPIC model would be presented. The model would then be implemented in python.

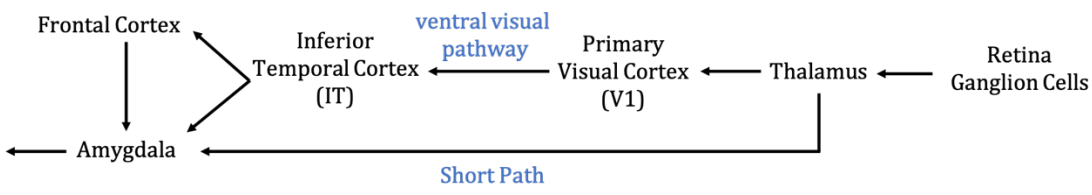
3.1 Emotion in the brain: The Limbic System

The components in a neural pathway collectively constitute a system. In the history of neurobiology, there have been debates about whether or not there exists a system that process the experience of emotion, which later becomes **the limbic system**. The limbic system consists of the amygdala, hippocampus, cingulate cortex, hypothalamus, thalamus, and the neocortex, which contains the orbitofrontal cortex (OFC) (Fig. 2).



(Fig. 2) The components of the limbic system [5]

When a person sees an image, the visual information goes through the following pathway: the retina ganglion cells response to light and visual information enters the thalamus and follows from the primary visual cortex (V1) to the inferior temporal cortex (IT), which forms the ventral visual pathway. The neurons in V1 response to low level features and semantic features are formed on the IT. This is the usual path that visual information takes. However, if the stimulus (image) presented is emotional or if it triggers emotion, this information is send to the **amygdala** in the limbic system as well. This pathway is shown in Fig. 3.



(Fig. 3) Pathway for visual information when emotional stimuli is presented.

Studies has shown that the human brain recalls emotional images more (ex: scary scenes) than it does to normal images [6]. This implies that not only emotional response is involved, but long-term "permanent" memory also plays a crucial role. The amygdala is associated with these two functions, interacting with the other components in the limbic system. The short path from the thalamus to amygdala allows quick response to stimuli. After receiving an emotional input, the amygdala interacts with the **orbitofrontal cortex (OFC)**, which evaluates the amygdala's emotional response and prevent it from being too easily triggered. Hippocampus also plays a role here, since it provides memory to the OFC to do the evaluation.

3.2 The "Brain Emotional Learning- Based Visual Pattern Recognizer" (BELPIC)

BELPIC is a model simulating the visual pathway and its interaction with the limbic system (Fig. 4). The stimulus is an image, whether emotional or not. The image is represented as a vector of n-dimension, $[S_1\ S_2\ ...\ S_n]^T$. To simplify the model, the networks in the thalamus and IT cortex are only considered as a path for visual information to go through. There are no weights in the thalamus or IT cortex since no information transformation are performed there. The amygdala receives input from the IT cortex and from the thalamus through a short path. This short path is for simulating the quick emotional response mentioned above. The weights in the amygdala are represented as a (n+1)-dimensional vector $[v_{1,1}\ v_{2,1}\ ...\ v_{n,1}\ v_{n+1,1}]^T$, and the weights in the OFC are represented as a n-dimensional vector $[w_{1,1}\ w_{2,1}\ ...\ w_{n,1}]^T$. Ea_1 is the emotional response of the amygdala while Eo_1 is the evaluation by the OFC. The amygdala interacts with OFC, which evaluates if the emotional response is appropriate. Hence, the total emotional response E_1 is the difference between Ea_1 and Eo_1. The equations are as follows:

$$Th = max\ (S_i) \tag{1}$$

$$Ea_1 = hardlim(\sum_{i=1}^{n} v_i \cdot S_i + v_{n+1} \cdot Th) \tag{2}$$

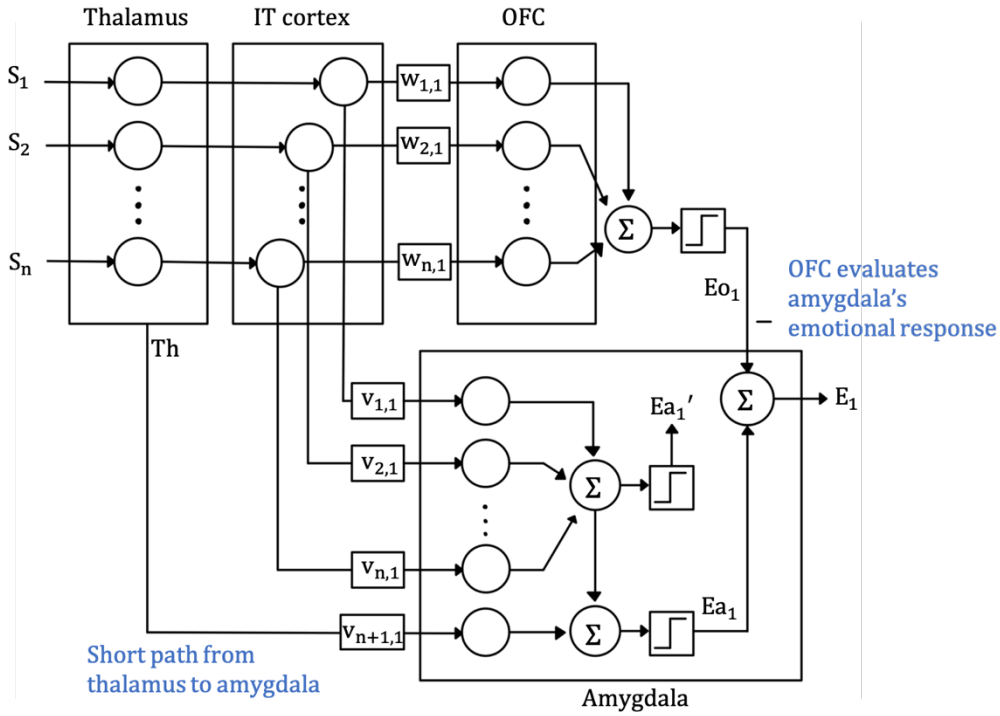
$$Eo_1 = hardlim(\sum_{i=1}^{n} w_i \cdot S_i) \tag{3}$$

$$Ea_1' = hardlim(\sum_{i=1}^{n} v_i \cdot S_i) \tag{4}$$

$$E_1 = Ea_1 - Eo_1 \tag{5}$$

, where $hardlim(x)$ is an activation function described as:

$$hardlim(x) = \begin{cases} 1, & if\ x > 0 \\ 0, & else \end{cases} \tag{6}$$



(Fig. 4) The structure of the BELPIC model

The goal of the model is to **correctly response to emotional images and normal images**. If an emotional image is shown, we expect the model to output 1 at E_1, simulating the scenario that the amygdala is activated when triggered by emotions. Otherwise, the model is expected to output 0 at E_1, implying that the amygdala has not been activated by a normal image.

To ensure the correctness of the output, the model would have to undergo a training phase. During the training phase, training images are shown to the model, and weights of the amygdala and OFC are updated in order to learn the correct response. The training rules are as follows:

$$v_i^{t+1} = (1-\gamma) \cdot v_i^{t} + \alpha \cdot max(R_1 - Ea_1, 0) \cdot S_i \quad (7)$$

$$v_{n+1}^{t+1} = (1-\gamma) \cdot v_n^{t} + \alpha \cdot max(R_1 - Ea_1, 0) \cdot Th \quad (8)$$

$$w_i^{t+1} = (1-\gamma) \cdot w_i^{t} + \beta \cdot R_0 \cdot S_i \quad (9)$$

, where

$$R_0 = \begin{cases} \max(Ea_1' - R_1, 0) - Eo_1 & ,if\ R_1 = 1 \\ \max(Ea_1' - Eo_1, 0) & ,if\ R_1 = 0 \end{cases} \quad (10)$$

α and β are the learning rates of the amygdala and OFC, respectively. After each iteration, the weights are all multiplied by $(1-\gamma)$ so as to simulate the "forgetting process" – if a stimuli was presented a long time ago, the human brain would eventually forget. The "learning process" is simulated by the term multiplied by the learning rates, where R_1 is the reinforcement term. If the image shown is emotional, R_1 is set to 1 to increase the response of the amygdala (reinforce the emotional memory). Otherwise, R_1 is set to 0. Note that in the learning process, the weights monotonically increase, while the only way to reduce them is through the forgetting process. This is because in the human brain, forgetting does not occur by reversing the acquisition pathway [7]. Hence, the two processes should be described independently.

3.2 Implement BELPIC in python

To test the proposed model, I downloaded a facial expression image dataset (https://www.kaggle.com/c/ml2019spring-hw3/data). The dataset contains images (size 48x48) of human faces and the labeled facial expression, which would either be one of the basic emotions (angry, disgust, fear, happy, sad, and surprise) or neutral (unemotional images). Since the model mentioned above is only capable of telling the difference between emotional and non-emotional images, only the images labeled "angry" and "neutral" are used. In total, there are 500 images in the dataset, 256 images labeled "angry" and 244 images labeled "neutral".

When an emotional image (angry) is represented, the reinforcement term R_1 is set to 1. The weights of the amygdala and OFC was initialized randomly, sampled from a uniform distribution between 0 and 1. α and β were set to 0.5 and 0.3 respectively.

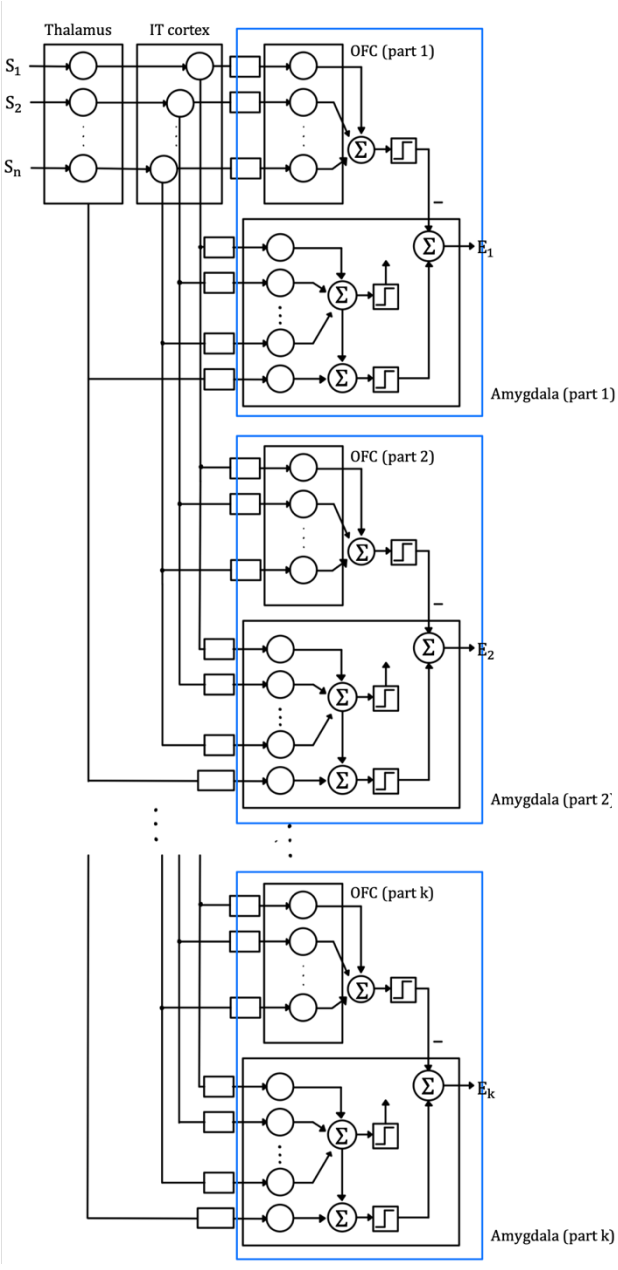
4. Results: The BELPIC Model

4.1 Forgetting Rate

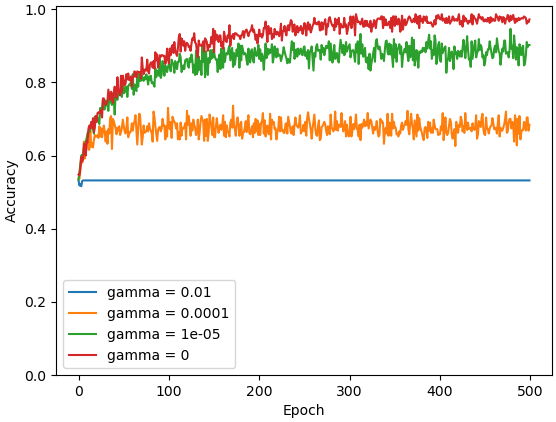
The model learns better if the forgetting rate (γ) is as small as possible (Fig. 5). When the forgetting rate is 0.01 (blue line), the model barely learns anything with the accuracy stuck at 50% as if the response is random. This is consistent with our intuition: a person with a better memory learns better. If the forgetting rate is set to 0 (red line), the model can yield a nearly perfect accuracy. However, this is impossible for human brains. Thus, a more plausible parameter selection could be the green line, when the forgetting rate is 0.00001.

4.2 Recognition of Multiple Emotions

So far, the BELFIC model is capable of responding to emotional images but fails to recognize which type of emotion expression is presented. We can achieve this by splitting the amygdala and OFC into parts, each part responsible for a basic emotion (Fig. 6). Assume the first part is responsible for anger. When an angry image is shown, then the output E_1 would be 1, while other outputs $E_2, \ldots, E_k$ are 0 since their corresponding amygdala is not activated. For each part of the amygdala and OFC, the structure remains the same and the weights are independent of other parts. To test the feasibility of this model, it is also implemented in python. The amygdala and OFC are split into seven parts. Each part is respectively responsible for angry, disgust, fear, happy, sad, surprise, and neutral. The training results are shown in Fig.7.

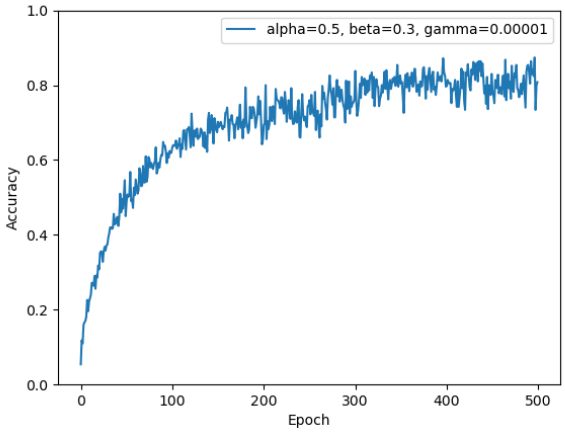


(Fig. 6) The extended version of BELPIC.



(Fig. 5)

Emotion response accuracy with different forgetting rates. The model outputs 1 for emotional (angry) images and 0 to neutral images.



(Fig. 7)

The training process for the extended version of BELPIC. The model is able to recognize the different emotions from the images and has an accuracy of approximately 80% after 500 epochs. For one epoch, 500 images of different emotions are shown to the model.

5. Discussion and Future Work

Both models are able to attain a high enough accuracy and produce the correct emotional response to the training images when the forgetting rate is small enough. As the forgetting rate increases, both models fail to learn and are not able to response correctly. In the first model that distinguishes anger from neutral, 48*48*2=4608 weights are used. In the second model which distinguishes between different emotions, 48*48*2*7=32256 weights are used. It took less than 5 minutes to finish the training phase for both models. This is a rather small amount of computational time and space compared to a CNN model with similar achievements. Hence, the model has the potential to perform better than the CNN model given very limited time and computational resources.

Although this model produces well enough results and attempts to imitate the functions and interactions between the components in the limbic system, the underlying structure can still be improved by enhancing its fidelity to the actual amygdala structure. For example, one can simulate the emotion recognition process with neurons and connections that do exist in the amygdala and OFC. The impact of hippocampus, thalamus, and IT cortex may also be considered. Further understanding about the actual structure of the limbic system is required to do so.

6. References

[1] Gupta, R. et al. (2012) The amygdala and decision making

[2] Ong, D. et al. (2019) Computational Models of Emotion Inference in Theory of Mind: A Review and Roadmap

[3] Ekman, P. (1999) Basic Emotions

[4] Lotfi, E. (2014) A Neural Basis Computational Model of Emotional Brain for Online Visual Object Recognition

[5] Bear, Mark F. et al. (2016) Neuroscience: Exploring the brain (4^{th} edition)

[6] Cahill, L. (1996) Amygdala activity at encoding correlated with long-term, free recall of emotional information

[7] Davis, Ronald L. (2017) The Biology of Forgetting – A Perspective