E-Bot: A Facial Recognition Based Human-Robot Emotion Detection System

Rahul Sridhar
Indian Institute of Technology
Dhanbad, Jharkhand, India
rahul.sridhar97@yahoo.com

Haiying Wang
Ulster University
Newtownabbey, Antrim, UK
hy.wang@ulster.ac.uk

Patrick McAllister
Ulster University
Newtownabbey, Antrim, UK
McAllister-P2@ulster.ac.uk

Huiru Zheng*
Ulster University
Newtownabbey, Antrim, UK
**h.zheng@ulster.ac.uk*

There are many Emotion Detection Systems which can understand emotion but are only present for monitoring purposes. Emotional Robots currently available predict emotion with low accuracy. Thus there is a requirement for Emotion Detection Robot which can predict emotion of user with high accuracy. The robot should converse with the user based on their emotion. It should be a companion for the user. In this paper we present the development of the E-Bot system, which enables human-robot interaction based on emotion detected from facial recognition. A mobile app is developed to control the robot and guide the chat. The Google Cloud Vision API and Pre-trained Facial Expression Algorithm are explored and the prediction accuracy, sensitivity and specificity of these two approaches are summarised in this paper. Results show that the proposed E-Bot system could be applied to provide affective care for people living alone at home.

Affective Computing, Emotion Detection, Human-robot, Chatting Robotics

1. INTRODUCTION

For a robot to converse like humans it is essential that it understands how the human feel. Affective computing or emotional Artificial Intelligence is the study and development of systems and devices that can recognize, interpret, process, and simulate human affects/emotion (Fingar, 2015). It combines engineering and computer science with psychology, cognitive science, neuroscience, sociology, education, psychophysiology, value-centered design, ethics, and more (Fingar, 2015). The aim is to restore a balance between emotion and cognition in the design of technologies for addressing human needs (Fingar, 2015).

Much research has been completed in researching and developing emotion detection systems to detect the emotion of the user based on the data collected by sensors and cameras installed to monitor the user. There are robots which can converse with the user based on their emotion but they have a poor accuracy of predicting the emotion. Thus, more research is needed in applying affective computing technologies with chatting robots to personalise conversation with the user. Previous research has shown that robots which can understand the emotion of user could be used in the following ways: (i) change style of teaching to keep the student focused (Firth, 2012), (ii) therapy (Obias, 2015), (iii) companion for patients (Obias, 2015), (iv) counselling to give advise based on client's emotional state (Molteni, 2017), (v) in healthcare to find how the patients are feeling about treatment (Jaiprakash et al., 2016), (vi) companion for people suffering with dementia (Broadbent, 2017).

Our aim is to build an emotion detection chatting robot which can come when the user calls it and have a conversation based on their emotion. It could be like a companion. In this work, we propose a chatting robot, E-Bot, which is able to interpret user's emotion and begin the conversation based on the emotion using visual sensors and speaker. E-Bot is the culmination of different hardware and software technologies that is able to detect the emotion of the users using a camera (facial expressions) and converse with the user based on the emotion (text to speech conversion). The proposed EBot system could be a replacement to emotion detection systems like 'affective computer tutor'(Picard, 1995) and 'intelligent room'(Hirish et al., 1999) as well as chat bots. In the paper "Experience from the operation of the Pepper humanoid robots" the authors (Gardecki et al., 2017) mention that the state of the art emotion detection and interaction robot Pepper predicts emotion of the user with a very low accuracy. Thus

Proceedings of British HCI 2018.
Belfast, UK.

there is a need for a robot which can predict accurately the emotion of the user and interact with them based on the how they feel.

2. PROPOSED APPROACH

In our approach we use an Empathy Bot to interact with the user based on their emotion. To detect the emotion we use two approaches – Google Cloud Vision API and Facial Expression Detection (FED) algorithm by Serengil (2017). Both of these approaches use convolutional neural network trained on labelled dataset. Google Cloud Vision API was used because it enables to understand the content of an image by encapsulating powerful machine learning models in an easy to use REST API. The API takes an input JSON response containing the image. It then quickly detects the faces in an image, predicts the emotion and returns an output JSON response which contains the emotion. One of the major drawbacks of using Google Cloud Vision API for predicting emotion is that it requires sending the image to Google server. If the application involves sensitive information for eg. monitoring patients with Dementia then this approach cannot be used. That is why we used another approach wherein we send the image to the remote server on which we ran the FED algorithm by Serengil (2017).

Earlier approaches to detect emotions included (1) Defining expression in terms of muscle action system to categorize the physical expression of emotions (Ekman, 1972 y 1999). (2) Facial Action Coding Systems (FACS) (Ekman, 1972 y 1999; Facial Action Coding System(FACS) and FACS Manual, 2011). (3) Multimodal recognition, e.g. Facial expression and speech prosody (Clever, 2011) to provide robust estimation of the subject's emotional state (Soft Computing, 2011). (4) Optical flow (Ekman, 1972 y 1999; Facial Action Coding System(FACS) and FACS Manual, 2011). (5) Hidden Markov model. (6) Artificial Neural Network(ANN) processing or active appearance model. (7) for feature extraction, many well-known handcrafted feature, such as HoG, LBP, distance and angle relation between landmarks are used and the pre-trained classifiers, such as SVM, AdaBoost, and random forest, are also used for facial expression recognition based on the extracted features (Suk, 2014).

2.1 E-BOT SYSTEM HARDWARE OVERVIEW

We have used a robot for the purpose of emotion detection of the user and to start a conversation with them. Figure 1 shows the components of the robot. The Raspberry Pi 3 Model B+ acts as the brain of the bot. We use Raspberry Pi because it is credit-card sized computer, low cost and is available anywhere in the world. The Raspberry Pi Camera Module v2 is used to take pictures. A speaker is connected to the Raspberry Pi for allowing the robot to speak with the user. The Raspberry Pi is stacked on top by a GoPiGo3 board. The communication between the two occurs over SPI interface. The motors for the wheels of the bot and the Distance sensor are connected to the GoPiGo3. The Raspberry Pi communicates to the motors and the sensor via GoPiGo3 board.

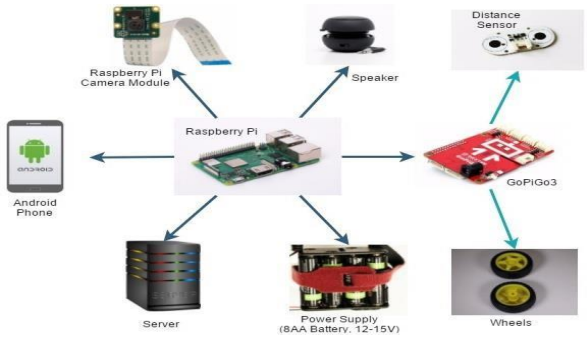

Figure 1: *Mapping of the different parts of entire system.*

2.2 E-BOT SYSTEM SOFTWARE OVERVIEW

The Python main libraries used in the Raspberry Pi are easygopigo3 and Flask. The easygopigo3 library is responsible for moving the robot left, right, forward and backward and to find the distance. The Flask library is used to send a GET requests and send files. In the remote server the Python libraries used are Tensorflow and Keras which is a deep learning framework with Tensorflow running on its backend. The pre-trained model is implemented using Keras.

The E-Bot interface was developed using Java for deployment on Android-based smartphones. The E-Bot consists of 5 activities. An activity represents the presentation layer of an Android application i.e. a screen that the user sees. The first activity is used to get the name of the user. The second activity displays connection status between the Raspberry Pi and the Android Phone. The third activity displays the internet connection status for the Raspberry Pi and asks permission from the user if they would like the robot to come for an interaction. If the user enters 'Yes' it directs the user to the fourth activity. If user enters 'No' it directs the user to the fifth activity. This confirms if the user is sure about their decision. If 'Yes' directs the user to the third activity. If 'No', directs the user to the fourth activity.

When the robot reaches close enough to user the fourth activity displays options to rotate the bot left/right by 90 degrees if necessary so that it can face to the user and 'Start a Conversation. If the user selects the button 'Start a Conversation' the bot asks the user to look at it by saying 'Look at me. Let's have a chat.'. It then gets the image of the user and then starts a conversation based on the emotion of the user and displays the conversation on the app. After this the app displays the image of the user used to detect the emotion. After the conversation is over the robot moves back. To detect the emotion of the user

we used two approaches - Google Cloud Vision API and FED algorithm (Serengil, 2017). The FED algorithm was trained on FEC2013 dataset and achieved a training accuracy of 92.05% and test accuracy of 57%. Its test accuracy might not seem decent but is better than winning model of Kaggle challenge which got 34% accuracy (Serengil, 2017). The pre-trained model consists of 3 convolutional layer and three fully connected layer. The activation except for the output layer is 'relu'. For the final layer a softmax activation is used. The model is run on Keras framework.

A server is run in a remote location to store the image of the user. The communication between the Raspberry Pi and the server is based on Client-Server Model. Since Raspberry Pi has Flask server running on it when the Server wants to send a GET request the Raspberry Pi can act as a server and the remote server then becomes a client.

For communication between the Raspberry Pi and the Android Phone a Client-Server model was followed with the Raspberry Pi as the Server and the Android Phone as the client. We used a Flask server on the Raspberry Pi.

The flask server runs a python script as soon as the Raspberry Pi boots up. The robot does a specific task on tapping the corresponding button in the app with the help of GET request. This request is sent to the Flask server when the button is pressed on the app. Each task performed by the E-bot is a result of the execution of the corresponding function which is a part of the python script running on the Raspberry Pi. The function to be executed is determined by the URL sent with the GET request.

The entire process consists of the following steps:

- Enter user name;
- Checking for connection between the Raspberry Pi and the Android Phone;
- Check Internet connection for Raspberry Pi;
- Getting permission from the user if they would like the bot to come for an interaction;
- Calculating distance from user and stop on reaching close enough;
- Guide the E-Bot to 'Turn Left', 'Turn Right' or 'Start a conversation';
- Take a picture and detect the emotion (explained in section 2.3.1);
- Start a conversation based on the emotion of the user (explained in section 2.3.2); and
- Display the image to the user

2.3.1 Take a picture

After reaching close enough to the user the robot takes a picture of the user with the help of Pi Camera. It then redirects to the server. The server stores the image from Raspberry Pi. This image is used by the pretrained model to predict the emotion which is then sent to the Raspberry Pi. The image is temporarily stored in the Raspberry Pi to display it on the app. The dataflow is as shown in Figure 2.

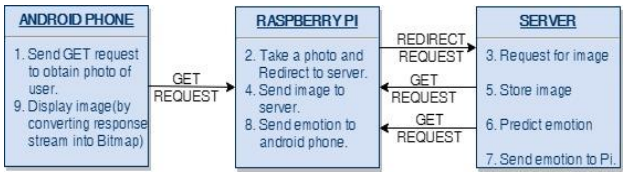

Figure 2: *Data Flow of the processes taking place in Android Phone, Raspberry Pi and Server for taking a picture, storing in server and displaying on the app.*

2.3.2 Begin a conversation based on the emotion of the user

One approach based on pretrained model to obtain emotion is discussed above. To obtain emotion using Google Cloud Vision API an input JSON response containing the image is passed to the API. The robot begins the conversation based on the emotion obtained as illustrated in Figure 3.

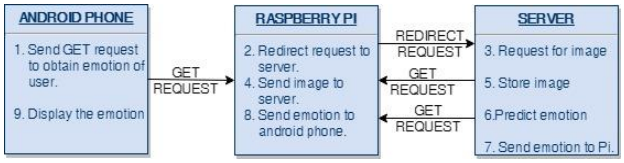

Figure 3: *Data Flow of the processes taking place in Android Phone and Raspberry Pi to find emotion from the picture using Google Cloud Vision API.*

Figure 4 shows the screenshot of the app after the conversation is over. The app displays the conversation with the user. It can be seen that the bot recognizes happy emotion of the user and replies accordingly.

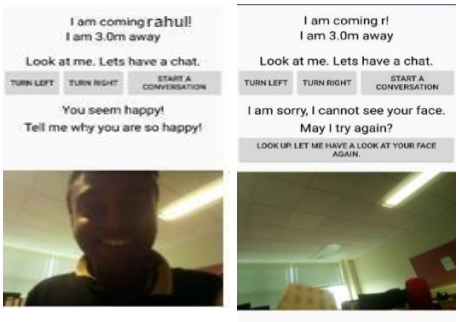

Figure 4: *Activity 4 to display the distance of the bot from the user, ask permission from the user about turn left, turn right or start a conversation and display emotion and image. If emotion is not detectable asking user to try again.*

2.4 TESTING APPROACH

For testing purpose, we asked five participants to enact six universal emotions anger, fear, disgust, surprise, joy and sadness (Ekman, 1971). We then recorded the emotion predicted by the bot for the two

approaches – Google Cloud Vision API and pretrained FED model. This was used to find the prediction accuracy, specificity and sensitivity of the algorithm to compare and analysis the performance of the two algorithms.

3. RESULTS & DISCUSSION

We tested our two approaches on five participants. Results are summarised in Table 1 and Table 2.

Table 1 shows the prediction accuracy, sensitivity and specificity of Google Cloud Vision API. The Google Cloud Vision API detects the emotions joy, sad, disgust and surprise with 80% accuracy. It detects the emotion fear with 40% accuracy and does not predict any anger emotion correctly. Thus any emotion which involves anger would pull down the accuracy of this approach. The overall test accuracy on the five people is 64.28%. The overall sensitivity and specificity of the algorithm in percentage is 60% and 78.67% respectively.

Table 1: *Prediction accuracy, sensitivity and specificity USING Google Cloud Vision API for five people with six different emotions.*

Emotion	Google Cloud Vision API		
	Accuracy	*Sensitivity*	*Specificity*
ANGER	*0%*	*0%*	*100%*
DISGUST	*80%*	*66.67%*	*66.67%*
FEAR	*40%*	*66.67%*	*66.67%*
JOY	*80%*	*80%*	*76%*
SAD	*80%*	*66.67%*	*66.67%*
SURPRISE	*80%*	*80%*	*96%*

Table 2: *Prediction accuracy, sensitivity and specificity using the FED algorithm by Serengil (2017) for five people with six different emotions.*

Emotion	FED Algorithm		
	Accuracy	*Sensitivity*	*Specificity*
ANGER	*20%*	*20%*	*92%*
DISGUST	*0%*	*0%*	*100%*
FEAR	*60%*	*60%*	*68%*
JOY	*40%*	*20%*	*92%*
SAD	*0%*	*0%*	*88%*
SURPRISE	*0%*	*0%*	*100%*

Table 2 shows the prediction accuracy, sensitivity and specificity of Facial Emotion Detection algorithm by Sefik Ilkin Serengil. The Facial Emotion Detection algorithm by Sefik Ilkin Serengil has a disappointing performance on real world example. It predicts the emotions fear with accuracy of 60%, joy with accuracy of 40% and anger with an accuracy of 20% which is relatively better than other emotions. It does not predict disgust, sad and surprise emotion correctly for any person. It has an overall accuracy of 20%. The overall sensitivity and specificity of the algorithm in percentage is 16.67% and 90% respectively.

4. CONCLUSION

In this study, we propose the development of Empathy Bot in which a number of sensors are utilised to determine emotion of an individual. It is envisaged that this system would be used to help people living with dementia by providing emotional support. Empathy bot combines voice analysis and deep learning approaches to interact with the user and initial experiments for emotion recognition were conducted using 2 methods; Google Cloud Vision API and FED Algorithm. Results from Google Cloud Vision API were promising and achieved 64.28% accuracy. In regards to future, the Google Cloud Vision API or Facial Expression algorithm would be replaced with our own CNN algorithm. Both the above approaches used in this study is a CNN however more work needs completing in utilising state-of-the-art CNN architecture such as ResNet-152 or InceptionResNet for image emotion detection. Research will also be completed in determining emotion in images and video using state-of-the-art CNN architecture. But to get a good accuracy it is important to predict the emotion from a video i.e. user's behaviour for a short span of time. This would help us in achieving higher accuracy.

In future iterations of E-Bot, further development will focus on refining and personalising the conversation after emotion have been classified and validation using a large cohort. Currently, the E-bot only supports Android Devices. Future work will focus on developing the app for other operating systems. From the hardware perspective we will be working on ways to reduce the cost of the bot. This would give the user a low cost accurate emotion detection chatting robot.

Acknowledgement

This research is partially supported by H2020 RISE SenceCare project (H2020-MSCA-RISE-2015) funded by the European commission. We thank Prof. Mike McTear for the useful discussion and feedback on the project.

5. REFERENCES

Fingar, P. (2015) The Cognitive Computing Era: Affective Computing. https://bpm.com/bpmtoday/blogs/997-affective-computing (latest access on May 25, 2018).

Broadbent, E. (2017) Can Robotics Help People With Dementia in the Community?

https://www.psychologytoday.com/gb/blog/health-happiness-and-robots/201707/can-robots-help-people-dementia-in-the-community (latest access on May 29, 2018).

Jaiprakash, A., Roberts J and Crawford R. (2016) Robots in health care could lead to a doctorless hospital. https://theconversation.com/robots-in-health-care-could-lead-to-a-doctorless-hospital-54316 (latest access on May 29, 2018).

Molteni, M. (2017) The Chatbot Therapist will see you now. https://www.wired.com/2017/06/facebook-messenger-woebot-chatbot-therapist/ (latest access on May 26, 2018).

Obias, R. (2015) 10 Therapy Robots Designed to Help Humans. http://mentalfloss.com/article/71987/10-therapy-robots-designed-help-humans (latest access on May 27, 2018).

Firth, N. (2012) Mind-reading robot teachers keep students focused. https://www.newscientist.com/article/mg21428665-500-mind-reading-robot-teachers-keep-students-focused/(latest access on May 28, 2018).

Serengil, S. I. (Last Updated 2017) Tensorflow 101: Introduction to Deep Learning. https://github.com/serengil/tensorflow-101 (latest access on May 26, 2018).

Ekman, P. (1972) Universals and Cultural Differences in Facial Expression of Emotion", In Cole, J. Nebraska Symposium on Motivation. Lincoln, Nebraska: University of Nebraska Press.

Facial Action Coding System (FACS) and the FACS Manual. (2011): A Human Face. N.p, n.d Web. 21 March. 2011

Clever. (2011). Algorithms. "Bacterial Foraging Optimization Algorithm – Swarm Algorithms – Clever Algorithms." Clever Algorithms. N.p., n.d. Web. 21/03/2011.

Soft Computing. N.p., n.d. (2011). "Soft Computing." www.softcomputing.net/bfoa-chapter.pdf, Mar, 2011.

Suk, M. and Prabhakaran, B. Real-time mobile facial expression recognition system—A case study. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 24–27 June 2014; pp. 132–137.

Ekman, P. and Friesen, W. V. (1971). Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, 17(2), 124129.

Gardecki, A. and Podpora, M. (2017). Experience from the operation of the Pepper humainoid robots. Progress in Applied Electrical Engineering (PAEE), 2017, 81-86.