Poster: Enhanced Chatting Based on Multimodal Emotion Estimation

Luyao Chong
School of Software and BNRist, Tsinghua University
chongluyao@gmail.com

Junchen Guo
School of Software and BNRist, Tsinghua University
gjc16@mails.tsinghua.edu.cn

Jinming Li
School of Software and BNRist, Tsinghua University
jm-li15@mails.tsinghua.edu.cn

Haozhen Liu
School of Software and BNRist, Tsinghua University
liu-hz16@mails.tsinghua.edu.cn

Meng Jin
School of Software and BNRist, Tsinghua University
mengj@mail.tsinghua.edu.cn

Yuan He
School of Software and BNRist, Tsinghua University
heyuan@tsinghua.edu.cn

Abstract

Currently online chatting applications only allow users communicate with pure text messages. These pure text messages can hardly convey emotions and ambiguities of language and words may lead to unnecessary misunderstandings and troubles for users. We proposed a new online chatting application named "EmotionChat", which allowed users know about the pleasure degree of each other during chatting. We analyzed the pleasure degree of the user using both facial expressions and message contents, and executed all of the process on the mobile devices. We aimed to reduce the time delay and energy consumption of the algorithm, so that we could analyze the pleasure degree real-time on the mobile device.

1 Introduction

Communication is an important way for people to share knowledge and enhance mutual understanding. In the process of communication, people express their opinions or feelings through language and words. However, language and words often have certain ambiguities [6]. In different contexts, the same language and words can express different contents. In face-to-face communication, people can infer the emotional state of the other side through his facial expressions and voice intonation, so as to better understand what he wants to express and avoid ambiguity. However, in online chatting, people only communicate through words. The lack of facial expressions and voice intonation leads people to be ignorant of the emotional state of the chat object. Therefore, the ambiguity of the text brings obstacles to communication and increases unnecessary misunderstandings and troubles [7].

We proposed "EmotionChat", a mobile application which can display the pleasure degree of the chat object in real time during the chat process, so that people can understand each other's emotional state and avoid unnecessary misunderstanding and trouble. Although emotions can be seen as personal privacy, there are still many online chatting situations where people want to share their emotions. For example, we want to share our joy with our lovers or family members, or we want our friends understand our sorrow and comfort us. We used people's facial expressions and text content to infer the emotional state of both sides of the chat, and attached it to the message to send to the other side. Compared to letting users input their emotions themselves, our approach could analyze emotions automatically without bothering users. Besides, the emotion achieved from our application has higher confidence because it is from facial expression instead of user input. In the process of implementing this app, there were several major challenges:

- How to effectively combine the information of facial expressions and text content to give a precise estimation of the pleasure degree.
- How to quickly estimate the pleasure degree to ensure the real-time nature of the chat.
- How to reduce energy consumption as much as possible.

We used deep learning methods to process facial expressions and textual content, and combined the results to achieve the final pleasure degree. Unlike previous works [5, 1], we avoided transferring pictures and text to the server for processing, which saved network transmission time. We used a lightweight model to speed up processing and reduce energy consumption. In addition, we leveraged the user's input time to process image information to further speed up and further reduced energy consumption by shortening the camera's turn-on time.

International Conference on Embedded Wireless Systems and Networks (EWSN) 2019
25–27 February, Beijing, China

ISBN: 978-0-9949886-3-8

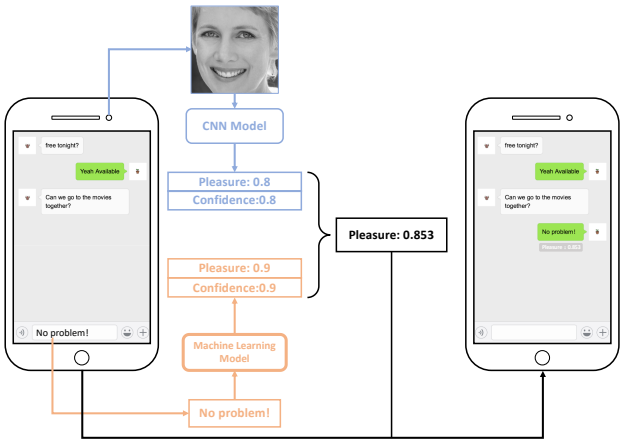

Figure 1. Overview of the architecture

2 Architecture

Figure 1 showed the architecture of EmotionChat. The interface of EmotionChat was very similar to a common chat application. But for each message, a hint of pleasure degree was attached below the message. The pleasure degree was inferred by the facial expression and message content when the user sends the message. When the app detects that the user was entering a message, it called the front camera to capture the user's current expression. The camera continued trying to detect the human face until it succeeded. After getting the face area, the front camera would be closed to save energy. And then the achieved face area would be fed to the deep learning model. The deep learning model would give the inference the pleasure degree. It usually takes a few seconds for a user to enter the message. We can process the facial expressions in this period, thereby reducing the time delay. And when the user clicked the button to send the message, the message content was achieved and submitted to the machine learning model for inference. After the inference using the facial expressions and the message content were completed separately, the final pleasure degree would be calculated based on the two results. And the final pleasure degree would be sent to the other side along with the message content.

3 Algorithm

Facial Expression Recognition

Recently, deep learning has been proved to deal with computer vision tasks effectively and efficiently. Convolutional neural networks also achieved a high accuracy on facial expression recognition [8, 3]. The development of hardware and deep neural network compression technology have made it possible to leverage convolutional neural networks on mobile devices. After obtaining the face area from the front camera using android API, we feed the face area into a vgg19 model trained with the fer2013 dataset [2]. The fer2013 dataset contained 28,709 examples as training set, 3,589 examples as public test set and 3,589 examples as private test set. The images are labeled as one of these categories: 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral. In order to match these seven emotion types to the pleasure degree, we use the recent research [4].

Message Content Analysis

In order to reduce the time delay, we use a light-weighted algorithm to analyze the message content to be sent. To deal with Chinese, we split the sentence into several words at first. Then we vectorize the words using the bag of words model. Finally, we use the naive bayes model to analyze the pleasure degree.

Multimodal information fusion After achieving the estimated pleasure degree from the facial expression and message content separately, we consider these two results comprehensively to get the final result. Assume that the pleasure degree achieved from the facial expression is D_{face}, and the confidence is C_{face}. While the pleasure degree achieved from the message content is $D_{message}$, and the confidence is $C_{message}$. Then the final pleasure degree is:

$$\frac{D_{face} * C_{face} + D_{message} * C_{message}}{C_{face} + C_{message}}$$

4 Discussion

We proposed "EmotionChat", an online chatting application which showed the pleasure degree of the user along with the message. It can help users know about each other's emotional state and avoid unnecessary misunderstanding and trouble. We used the facial expressions and message contents to analyze the pleasure degree. Our algorithm of information fusion is still naive. And we are focusing on finding out a better algorithm that can combine the information of the facial expressions and message contents effectively. Besides, there may be some privacy issues as we called the camera to get the user's facial expressions.

5 References

[1] J. Feijó Filho, T. Valle, and W. Prata. Non-verbal communications in mobile text chat: emotion-enhanced mobile chat. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, pages 443–446. ACM, 2014.

[2] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, et al. Challenges in representation learning: A report on three machine learning contests. In *International Conference on Neural Information Processing*, pages 117–124. Springer, 2013.

[3] H. Jung, S. Lee, J. Yim, S. Park, and J. Kim. Joint fine-tuning in deep neural networks for facial expression recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2983–2991, 2015.

[4] A. Landowska. Towards new mappings between emotion representation models. *Applied Sciences*, 8(2):274, 2018.

[5] C. Ma, A. Osherenko, H. Prendinger, and M. Ishizuka. A chat system based on emotion estimation from text and embodied conversational messengers. In *Active Media Technology, 2005.(AMT 2005). Proceedings of the 2005 International Conference on*, pages 546–548. IEEE, 2005.

[6] D. L. Nilsen. Ambiguity in natural language: An investigation of certain problems in its linguistic description, 1973.

[7] P. Resnik. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of artificial intelligence research*, 11:95–130, 1999.

[8] Z. Yu and C. Zhang. Image based static facial expression recognition with multiple deep network learning. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 435–442. ACM, 2015.