Developing Face Emotion Tracker for Quantitative Evaluation of Care Effects

Arashi Sako[1(✉)], Sachio Saiki[1], Masahide Nakamura[1], and Kiyoshi Yasuda[2]

[1] Graduate School of System Informatics Kobe University, 1-1 Rokkodai, Nada, Kobe, Japan
kosa-a@ws.cs.kobe-u.ac.jp, sachio@carp.kobe-u.ac.jp, masa-n@cs.kobe-u.ac.jp

[2] Chiba Rosai Hospital, 2-16 Tatsumidai-higashi, Ichihara, Japan
fwkk5911@mb.infoweb.ne.jp

Abstract. In 2017, the Japanese government declared to start "scientific long-term care" as a national policy. In the practice of scientific long-term care, it is essential to assess the quality and effect of care services, since the caregivers must know if the care was effective for the target person. Currently, however, the effect of the long-term care has been evaluated by subjective observation and/or the questionnaire. Hence, it is difficult to justify the quality and effect as such the evidence encouraged in the scientific long-term care. To cope with the challenge, this paper proposes a novel system Face Emotion Tracker (FET) that evaluates the effect of care as a transition of emotions of a person under care. The proposed system can produce real-time data quantifying emotions of the target person under care, which is more objective and fine-grained clinical data compared to the conventional manual assessment sheets. We also implement a prototype and conduct experiments using the prototype.

Keywords: Cognitive computing · Face recognition · Scientific care

1 Introduction

Currently, Japan has been facing a hyper-aging society. In 2025, 30.0% of the total people will be older than 65 years old [5]. As the increasing of elderly people, social problems are expected. The long-term care facilities are suffering from the chronical lack of caregivers and resources. There is also a big pressure for the cost of social security from the government. Therefore, it becomes more important to assign the limited care resources efficiently to appropriate targets.

In 2017, the Japanese government declared to start "scientific long-term care" as a national policy [6]. The scientific long-term care is a long-term nursing care whose effect must be justified by scientific evidence. It assumes to use quantitative and objective big data. By doing this, the scientific long-term care aims to provide cares optimized for individuals to support independent living. In the

V. G. Duffy (Ed.): DHM 2018, LNCS 10917, pp. 513–526, 2018.
https://doi.org/10.1007/978-3-319-91397-1_41

practice of the scientific care, it is essential to assess the quality and effect of long-term care services, since the caregivers must know if the care was effective for the target person.

Currently, however, the effect of the long-term care has been evaluated by subjective observation and/or the questionnaire. Both methods heavily rely on human subjective decisions based on an experience of caregivers. Furthermore, it takes caregivers and/or subjects a lot of time and effort to evaluate the effect of care. Hence, it is difficult to justify the quality and effect as such the evidence encouraged in the scientific long-term care.

To cope with the challenge, this paper proposes a novel system *Face Emotion Tracker* (FET, for short). FET evaluates the effect of care as a transition of emotions of a person under care. The emotions are captured from facial expressions using the latest cognitive computing technology. FET is comprised of four key features: (A1) face recognition, (A2) recording data, (A3) visualization of data, and (A4) quantification of effects.

In (A1), we periodically capture a face image of the target person under care, and send the image to Face Emotion API, which is one of cognitive cloud services. For a given face image, the API returns emotional values calculated from the facial expressions. For this, the emotion is characterized by the confidence of eight emotional attributes: (1) neutral, (2) happiness, (3) surprise, (4) sadness, (5) fear, (6) anger, (7) contempt, and (8) disgust. In (A2), a system operator records every treatment that a caregiver performs during the care. As a result, the system can record real-time data, describing what kind of emotion appeared by which care treatment, at when of the care session. In (A3), we visualize the measured emotions to see in real time or to look back afterwards. In (A4), We then propose metrics that quantify the quality of care, using the emotional values of happiness in care session.

In this paper, we implemented a prototype system of FET using Microsoft Azure Emotion API based on four key features. In practice, we conduct long-term care using the prototype for various people such as an elderly person and person with dementia, and examine the practical feasibility of FET.

2 Preliminaries

2.1 Long-Term Care in Practice

The long-term care refers to a nursing care for elderly people in a broader sense. It includes not only physical supports for daily living, but also activities to maintain healthy mental conditions and good quality of life. There are typical activities such as therapy, recreation, ensemble, and conversation. The purpose of these activities is to stabilize the emotion, to improve cooperativeness, and to encourage recollections.

Currently, the effect of the long-term care has been evaluated by subjective observation and/or the questionnaire. In the observation-based evaluation, a caregiver observes behaviors of the target person, and reads the emotions of the target person. For example, a caregiver evaluates the effect by observing, "Mr. Tanaka is smiling a lot when talking about movies" or "Mr. Sako seems to be sad when talking about old tales". In the questionnaire-based evaluation, a caregiver conducts a survey that asks the target person what were good and bad treatments, and what to be improved after care.

However, both methods heavily rely on human subjective decisions. Furthermore, it takes caregivers and/or subjects a lot of time and effort to evaluate the effect of care. It is therefore difficult to justify the quality of the care, as such the evidence encouraged in the scientific long-term care. In reality, the care programs at care facilities are often planned by subjective judgment and intuition, based on experience of caregivers.

2.2 Cognitive Computing

Cognitive computing is a computing paradigm where a system imitates the human brain and learns itself to derive answers [1]. It often refers to a smart system or algorithm that recognizes moods and emotions from non-numerical data, such as natural language, images, and sounds, and derives the values. Enabling technologies include natural language processing, emotion analysis, face analysis, and speech recognition. These can be used to implement a machine that supports human decision-making.

IBM Watson [3] is a famous platform of cognitive computing and has been used as a system to support clinicians such as discovering cancer and proposing medicine. In recent years, more and more companies have published cognitive computing APIs, which allows people and external applications to use the cognitive technology, easily and efficiently. Famous cognitive computing APIs include Google Cloud Vision API [2] and Microsoft Azure Emotion API [4].

3 Evaluation of Care Effect Using Face Emotion Tracker

3.1 Goal and Approach

The purpose of this research is to record objective fine-grained clinical data and quantitatively evaluate care for the practice of scientific long-term care. In order to objectively and quantitatively evaluate the quality and effect of care, we select what kind of data we need to acquire. For example, as care clinical data, there are data such as facial expression of caregivers and subjects, body movements, voices. There are also data on the environment that carries out care such as temperature and humidity. Since the influence on evaluation of the quality and effect of care depends on data, it is necessary to select data considering ease of acquired and handling, a noise of data.

In this research, we develop a system that evaluates the effect of care as a transition of emotions of a person under care. The reason why we use facial expression is that human facial expressions are universal in the world and it expresses most emotions. In addition, facial expression has abundant amount of information, and in the case of elderly care, facial expressions of the subjects can are easily acquired with a camera because the subjects see the caregivers. Also, in terms of system evaluation of care, it is easy to judge whether the evaluation performed by the system is correct by human observation. For these reasons, we use facial expressions in this research.

3.2 Face Emotion Tracker (FET)

In order to achieve objective assessment required by the scientific long-term care, we develop a novel system, *Face Emotion Tracker (FET)* in this paper.

FET evaluates the effect of care as a transition of emotions of a person under care. The emotions are captured from facial expressions using the latest cognitive computing technology. Figure 1 shows the main screen of FET. Before a caregiver initiates a care session, the caregiver inputs a title of session, a name of a target person, remarks in Session setting. For example, the title is "1st Dementia Counseling at Kobe Center Hospital", the target person is "Sako", the remarks are time, treatments, and target person"s characters.

Next, when the caregiver presses the Start button, FET periodically captures a face image of the target person with the web camera (as shown in lower left of Fig. 1), and sends the image to Face Emotion API. For a given face image, the API returns emotional values calculated from the facial features within the given image. In Recognition image (shown in lower right of Fig. 1), the recognized face

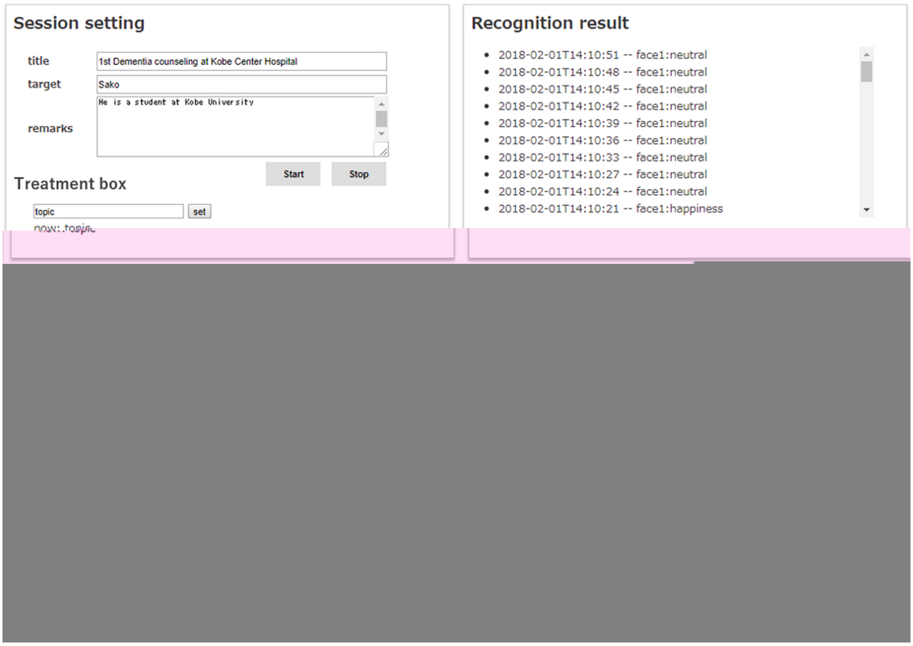

Fig. 1. Main screen of FET

Fig. 2. Visualization of data

is surrounded by a rectangle. In Recognition result (shown in upper right of Fig. 1), the attribute with the highest emotional value is displayed.

During the care, a system operator registers every treatment that the caregiver performs in the Treatment box. As a result, the system can record real-time data, describing what kind of emotion appeared in the target person, by which care treatment, at when of the care session.

Figure 2 shows the visualization feature of FET, visualizing the measured emotions using two graphs to see in real time or to look back afterwards. Furthermore, using the emotional values of happiness acquired during care, FET quantitatively evaluates care. We describe the details in the following subsections.

3.3 Face Recognition

In order to implement the face recognition feature of FET, we use Microsoft Azure Emotion API [4]. Figure 3 shows an example usage of Emotion API. For a given face image, the API returns coordinates of the recognized face, and emotional values calculated from the facial features. Emotion API tries to classify the given face image into eight emotion categories: (1) neutral, (2) happiness, (3) surprise, (4) sadness, (5) fear, (6) anger, (7) contempt, and (8) disgust. For each category, Emotion API returns a confidence representing how the given

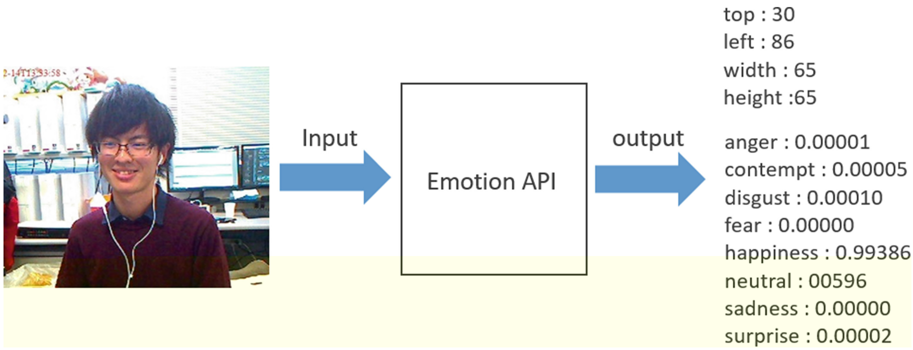

Fig. 3. Emotion API

face is likely to belong to the category. Therefore, we use the confidence as the quantitative indicator of the emotion.

Based on the coordinates, FET draws a rectangle indicating the recognized face. FET also displays an emotional attribute with the highest confidence in Recognition result. During the care session, FET executes the Face Recognition feature every 3 s.

3.4 Recording Data

As FET automatically measures the emotion values every 3 s, a system operator manually registers every treatment that the caregiver performs in Treatment box. For example, the system operator inputs like "singing", "ensemble", "conversation with Mr. Sako". As a result, FET can record real-time data, describing what kind of emotion appeared in the target person, in response to which care treatment, at when of the care session. FET stores the captured image, the attribute name, and emotional values, the treatment name, and the time in a relational database (RDB). Thus, it is possible to record more objective and fine-grained care clinical data, as compared with conventional manual evaluation.

3.5 Visualization of Data

As shown in Fig. 2, FET has a feature that visualizes the measured emotion data with two graphs. The explanation of each graph is as follows.

- G1: A bar chart visualizing the emotional values. The horizontal axis represents the eight emotion categories, while the vertical axis represents their confidence.
- G2: A line chart visualizing the time-series value of emotion. The horizontal axis presents time with treatment names, and the vertical axis plots the emotional values.

The visualization feature can be used in both during care and after care. In the case of during care, the two graphs change in real time. In G1, one can check what kind of emotion appeared in the target person immediately during care. In G2, one can check how the emotion is changing as the care session progresses. Thus, using two graphs, the caregiver can improve the treatment dynamically on the spot. In the case of after care, G1 displays the average value of each emotion throughout care or each treatment. In G2, a caregiver can see the transition of emotion throughout care or each treatment. Using two graphs, the caregiver can review the whole care and each treatment, and improve the treatments for the future care sessions.

3.6 Quantification of Effects

Using the emotion data captured by FET, we propose quantitative metrics, which evaluate the quality of care by the emotional value of happiness. The reason why we use happiness is that the long-term care aims to make elderly happy. On the other hand, a situation where the target person is always laughing is problematic, since there is a suspect of euphoria. Therefore, we assume that an ideal care session produces many happy emotions, where happiness and other expressions are alternatingly observed. Based on the assumption, we define the following two metrics:

$$E = \frac{1}{n}\sum_{i=1}^{n} P_i \tag{1}$$

$$V = \frac{1}{n}\sum_{i=1}^{n}(P_i - E)^2 \tag{2}$$

In the definition, n represents the number of facial images captured during a care session, and P_i represents the emotion value of happiness appeared in i-th image. Equation (1) represents the average value of happiness. Intuitively, E characterizes the degree of how the target person was happy throughout the session. Equation (2) represents the variance. Intuitively, V characterizes how dynamically the happiness appeared, distinguishing the happiness from the euphoria.

We can say that a care session A was better than B when the values of E and V for A are better than those of B. By using E and V, it is possible to quantitatively and objectively understand what treatment is suitable for the target person, what treatment is not suitable. As a result, a caregiver can find better treatment such as “Mr. Sako was very pleased when talking about Mr. Tanaka”. The caregiver may also find a symptom euphoria such as “Ms. Suzuki was always laughing at the time of the ensemble ”.

4 Implementation of Prototype

We have implemented a prototype system of FET based on the proposed method described in Sect. 3. Considering the portability of the system, the FET system

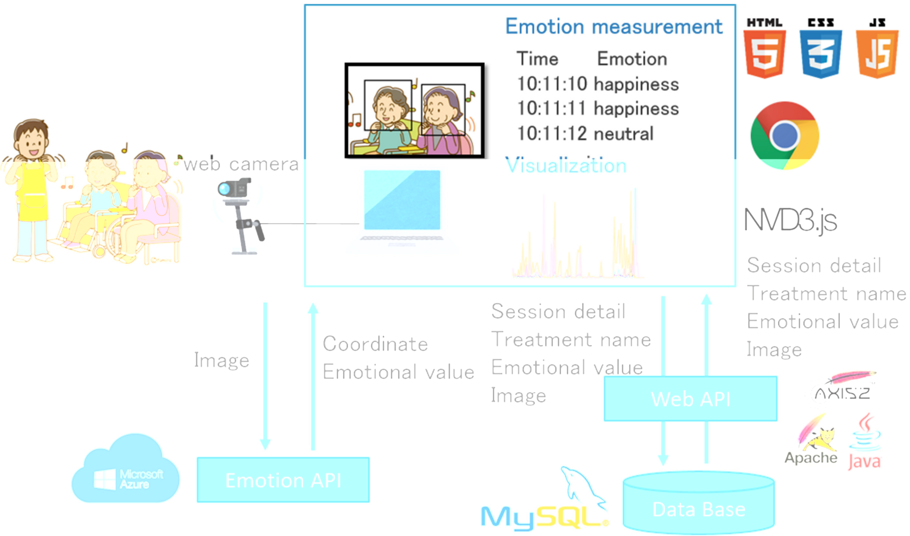

Fig. 4. System architecture of FET

was implemented as a Web application working on a Web browser. The user interface was implemented by HTML5, CSS3 and JavaScript. Just using a PC with a Web camera, the user can use FET from anywhere without installing any software.

The system architecture is shown in Fig. 4. During the care session, FET captures the face image and sends to the Emotion API of Microsoft Azure. When the Emotion API receives the image, it returns the coordinates of the face detected from the image, and eight values of the emotional attributes.

We also implemented Web API to record and retrieve care clinical data in the database, which was deployed on a server. By using the Web API, it is possible to record session detailed information, emotional value, face image data received from the Web service of the system in the database. In addition, these data are acquired using the Web API and used for visualization of data, evaluation of care.

Technologies used for the implementation are summarized as follows:

- Development language: Java1.7, HTML5, CSS3, JavaScript
- JavaScript library: JQuery 2.1.4, NVD3 1.8.1
- Web server: Apache Tomcat 7.0.69
- Web service framework: Apache Axis2 1.6.3
- Database: MySQL 5.7.18
- API: Microsoft Azure Emotion API

5 Experimental Evaluation

In this section, we conduct an experiment to evaluate the practical feasibility of the proposed system. The objective of the experiment is to see if the system can actually obtain quantitative emotion data from individual person under care. We also see whether or not the effect and quality of the care can be quantitatively evaluated by visualizing and analyzing the data.

All procedures used in the experiment were approved by the Ethical Committee of Graduate School of System Informatics, Kobe University (No. 29-04).

5.1 Experiment with Virtual Care Giver [7]

This experiment was performed together with another experiment of our project, where elderly people communicate with *Virtual Care Giver (VCG)* [7]. VCG is a system that supports elderly people at home by using the virtual agent (VA) technology. The VA is an animated chat-bot software with speech recognition and synthesis technologies. Through a PC screen, a user can talk to the VA who behaves like a human being.

VCG implements various sessions of communication care, using available Web resources including text, pictures, and movies. Each care session is extensively personalized to a person under care, considering his/her life history and circumstances [7]. The main goal of the experiment was to see how elderly people accept and communicate with the VCG.

Cooperated with the VCG project, our interest here is to evaluate the quality and effect of the care sessions provided by VCG, by using the developed FET system. In addition, we compare the results of different subjects, and see how the individual differences on communications and symptoms appear on the quantitative data. By doing this, we consider the validity and effectiveness of the evaluation method of the proposed system.

5.2 Method of Experiment

For the experiment, we asked an elderly day-care center to recruit elderly volunteers. As a result, five elderly people in the center kindly participated in the experiment. Let A, B, C, D, E denote the five subjects. Subject A was a man of 81 years old, and B, C, D, E were women of 99, 74, 84, 86 years old, respectively. Subject E was a person with heavy dementia.

Before the experiment, we asked the manager of the day-care center to collect questionnaire from five subjects, asking their personal information, including life history, circumstances, hobby, favorite music, and so on. The questionnaire was used to create personalized care scenarios of VCG.

VCG provides personalized communication care for each person. A session consists of multiple dialogue scenes, each of which talks about a certain topic personalized to the person under care. The dialogue scenes used in the experiments were: (1) self-introduction, (2) health check, (3) feature description

Fig. 5. Elderly communicating with Virtual Care Giver

(of the robot), (4) origin, (5) age of birth, (6) childhood, (7) work, (8) hobby, (9) favorite music, and (10) ending. Each session lasts about 20 min.

Figure 5 shows a scene in the experiment, where an elderly person is talking to VCG via a microphone. As seen in the picture, an operator of FET sits besides the subject. When a new dialogue scene starts, the operator manually inputs the scene name as a care treatment in FET. A Web camera placed on the display of VCG captures face images of the subject. It is connected to FET deployed in the laptop PC of the operator.

During each session, FET records the emotion values as clinical data, which is derived from the face image of the subject. FET also visualizes the values with charts, associated with the ten dialogue scenes.

5.3 Result of Experiment

Figure 6 shows a time-series graph of emotion values of Subject A, captured throughout the session. In the session, we observed that Subject A was moved to tears while he was singing his favorite music (i.e., the last part of the session). It appeared that the song might remind him of something nostalgic. It can be seen in Fig. 6 that the emotion value of sadness increased significantly in the scene of favorite music. Thus, FET captured well his transition of emotion from his face image.

Figure 7 shows the emotion values of Subject E. In the session, we observed that Subject E was gently smiling all the time during the session. This can be

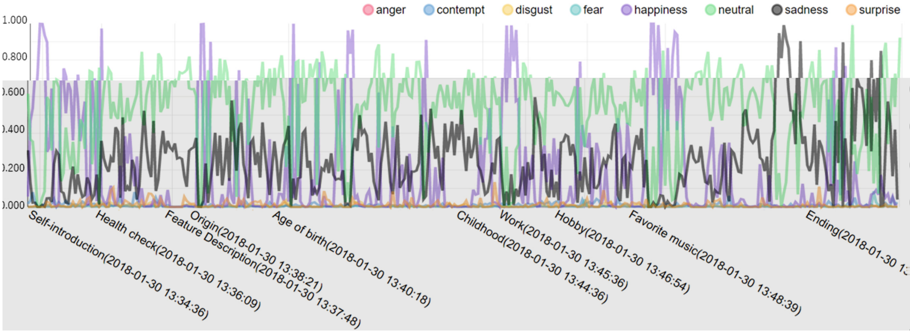

Fig. 6. Time-series graph of Subject A

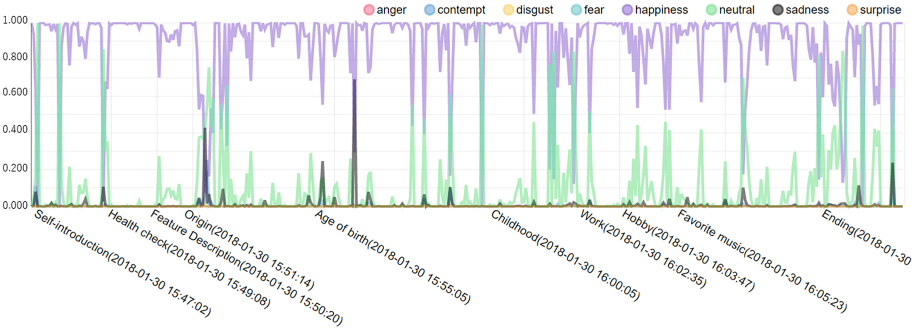

Fig. 7. Time-series graph of Subject E

seen in Fig. 7, where the emotion of happiness sticks to 1.0 most of the time. Thus, the proposed system was able to properly recognize the face, and record the care clinical data for the elderly people.

Figure 8 shows the average values of happiness, i.e. the metric E proposed in Sect. 3.6, for each dialogue scene. On the other hand, Fig. 9 shows the variance values of happiness i.e. the metric V proposed in Sect. 3.6, for each dialogue scene. These metrics give us interesting observation.

For example, as for Subject A, both metrics E and V take large values in the scene of Childhood significantly, compared to other scenes. This is justified by the fact that subject A talked about his childhood very cheerfully. Thus, talking about the childhood can be an effective communication care for Subject A.

5.4 Discussion

In addition to Subject A, the emotions of Subjects B and C dynamically changed in singing or listening to a song, which was observed in their time-series data (like the one in Fig. 6). In the interview after the experiments, they did not tell especially how the songs moved their minds. However, FET surely captured their emotions.

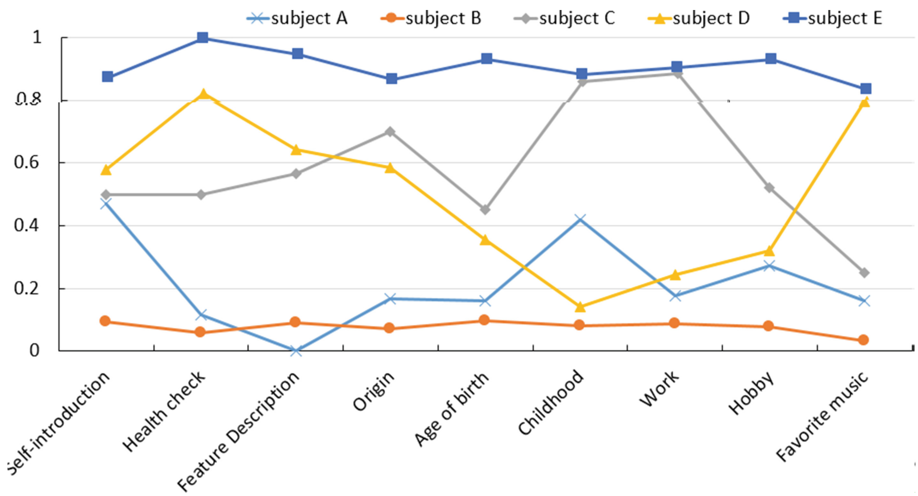

Fig. 8. Average of happiness (metric E)

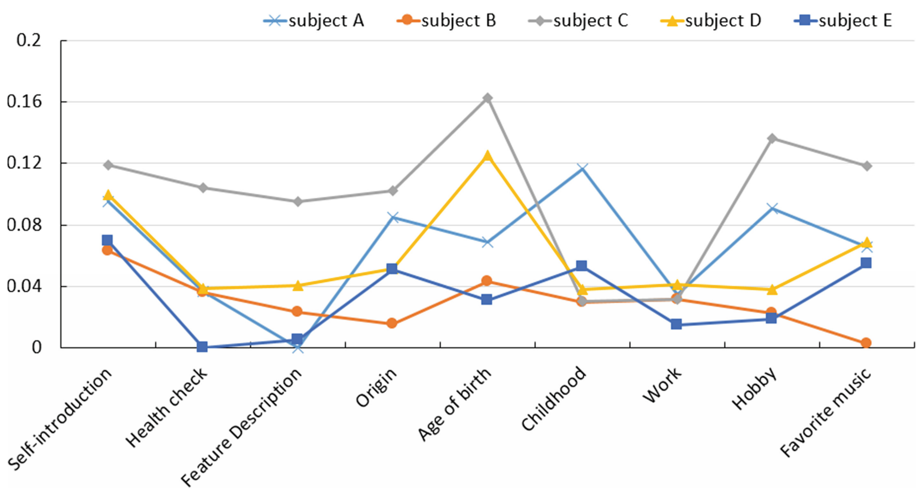

Fig. 9. Variance of happiness (metric V)

Indeed, in singing the song, emotions other than happiness were well observed, which could not quantified by the metrics E and V. Therefore, further investigation of effective metrics should be investigated. However, the clinical data acquired by the system can be used to provide evidence-based care considering emotional status of individuals.

As seen in Figs. 8 and 9, the empirical data shows that the happiness of Subject B was always low, and that her emotional ups and downs were very small, compared to other subjects. On the contrary, the care manager says that she enjoyed the communication with VCG well, which contradicted with the data. The reason is that Subject B is 99 years old, and her facial muscle has been weakened. Indeed, it is difficult even for general people to read her emotion from her facial expression only. Thus, this is the limitation of FET. In order to capture emotions from elderly people like her, we need to combine other features such as voice and postures.

Also, the empirical data shows that Subject E was always happy, and her emotional ups and downs were very small. As mentioned before, Subject E suffers from heavy dementia. Hence, the symptom of euphoria might be captured in the data. Since we have not yet consulted any doctors or medical professionals for this observation, this is yet an observation. However, the empirical data captured by FET would be used to characterize the euphoria and other disease. Further investigation will be left for our future work.

6 Conclusion

In this paper, we proposed a novel system, called Face Emotion Tracker (FET), to support the quantitative assessment of the quality of long-term care. Exploiting the latest cognitive computing technology, FET computes emotion values of a target person from facial images, and uses the values to define metrics characterizing the quality of care. We implemented a prototype system based on proposed methods and conducted experiments, and examined the practical feasibility of FET. Utilizing FET, we can contribute to the practice of scientific long-term care.

In our future work, we plan to combine other features related to emotions such as voice and postures. Introducing environmental data such as temperature and humidity is also interesting. Furthermore, we will propose other effective metrics quantifying the quality of care with other facial emotions.

Acknowledgements. This research was partially supported by the Japan Ministry of Education, Science, Sports, and Culture [Grant-in-Aid for Scientific Research (B) (16H02908, 15H02701), Grant-in-Aid for Scientific Research (A) (17H00731), Challenging Exploratory Research(15K12020)].

References

1. Cognitive Computing. https://en.wikipedia.org/wiki/Cognitive_computing
2. Google Cloud Vision API. https://cloud.google.com/vision/
3. IBM Watson. https://www.ibm.com/watson/
4. Microsoft Azure Emotion API. https://azure.microsoft.com/ja-jp/services/cognitive-services/emotion/

5. Cabinet Office, Government of Japan: Annual report on the aging society in 2017 (full version). http://www8.cao.go.jp/kourei/whitepaper/w-2017/html/zenbun/index.html
6. Headquarters for Japan's Economic Revitalization (No.26): Future investment strategy (2017). http://www.kantei.go.jp/jp/singi/keizaisaisei/dai26/siryou.pdf
7. Sakakibara, S., Saiki, S., Nakamura, M., Yasuda, K.: Generating personalized dialogue towards daily counseling system for home dementia care. In: Duffy, V.G. (ed.) DHM 2017. LNCS, vol. 10287, pp. 161–172. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58466-9_16