# A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation

Kyo-Joong Oh, DongKun Lee, ByungSoo Ko, Ho-Jin Choi

School of Computing, Korea Advanced Institute of Science and Technology (KAIST),
Daejeon, Republic of Korea
{aomaru, hagg30, kobiso, hojinc}@kaist.ac.kr

*Abstract*— There are early studies to attempt users for psychiatric counseling with chatbot. They lead to changes in drinking habit based on intervention approach via chat bot. The application does not con-sider the user's psychiatric status through the conversations, continuous user monitoring, and ethical judgment in the intervention. We contend that more accurate and continuous emotion recognition gives better satisfaction to users who need mental health care. In addition, appropriate clinical psychological response based on ethical responses is as well. We suggest a conversational service for psychiatric counseling that is adapted methodologies to understand counseling contents based on of high-level natural language understanding (NLU), and emotion recognition based on multi-modal approach. The methodologies enable continuous observation of emotional changes sensitively. In addition, the case-based counseling response model that combines ethical judgment model provides a suitable response to clinical psychiatric counseling.

*Keywords-conversational service; psychiatric counseling; mental healthcare; emotion recognition*

## I. INTRODUCTION

It has been a long research topic that machines recognize human emotions. Recently, many studies improves to recognize human emotions based on artificial intelligent (AI) methods. The studies train various emotion classification models from a lot of emotional-labeled data based on deep learning, such as convolution neural network [2], recurrent neural net-work [3], attention network [3, 4]. For the techniques are advanced, the training data are also diversified to image [2], video [2, 3], audio [4] and text [5]. Some studies [3, 5] com-bine the classification models using multi-modal classification based on hybrid approaches. The studies report significant results for emotion recognition. They are quite accurate to recognize the human emotions.

However, there are still few applications applied the emotion recognition techniques. Lately, some intelligent assistants service are in use; Apple Siri, Google now, Samsung S-Voice, and so on. These services respond to users' inputs, such as queries and voice, and suggest useful information for the users. However, the services just apply basic natural language processing (NLP), and key applications are shortfall yet. Consequently, the development of application for healthcare have studied with the intelligent assistant [6]. Already, some chatterbot and chat assistant are noticing favorably in commercial and customer service.

In this paper, we introduce an application of counseling chatbot, which provide conversational service for mental health care based on above emotion recognition methods and the chat assistant platform.

**Status** Lately, many patients undergo psychotherapy due to changes in awareness of psychiatric treatment. Although modern people have many mental illnesses, real rate of the diagnosis and treatment are still low. Since, it is not easy to be aware of mental illness of one selves is not easy generally. There is a lack of experts compared to demand, even the cost of consultations are expensive with the experts: psychiatrists, mental clinical counselors, and counseling psychologists. Therefore, an unfair problem of treatment opportunity arises. In order to solve the problem, such the unfair treatment opportunity, self-diagnosis and simple consultation services needs to encourage treatments by specialized medical institutions. Using the conversational service, anyone can easily accessible and receive practical treatments for mental counseling services.

**Example** The conversational service can provide personalized counselling service to individual head-to-head. It is important to resolve the isolation of the patients who have a mental dis-order such as depression and lethargy [7]. One-to-one conversation can resolve the isolation effectively. Personal dialogues can also operate efficiently when a user needs urgent interventions [1]. The service notify user's dangerous status when accidental mental disorder, such as panic and suicidal impulse. The suicidal requires immediate and quick response. It is possible to take steps that help to reduce suicidal impulse, and inform peoples who are around the user effectively. In addition, the service observe the mood swings continuously for users who have manic-depressive. Specific instructions pro-vides through dialogues among chat assistant and users. The conversational service for psychiatric counseling with the chat assistant has many benefits like these.

**Critique** In the early study [7], they have developed simple chat bot that provide interventions about subsequent behavior change for young adults' alcohol risk. They give the interventions based on three item questionnaires about drinking habits via alcohol consumption level and frequency of drinking by internationally well-known as AUDIT-C. The 17 participants are satisfied and perceived usefulness. However, the chat bot cannot adapt to long-term mental

disease, because there is no continuous observation. Ethical contemplation about interventions of chat bot considered it as well.

**Inherency** Linguistic interaction is the most important in psychiatric counseling. The interaction is one of typical psycho-logical treatment from Freud's psychoanalytic legend; it called insight-oriented psychotherapy. Patients can find a solution by increasing their understanding of the patient's own problems through verbal interaction. Therefore, users should not feel to talk to machine in the consultation, for to maximize the effect of the consultation. In addition, it is necessary high-level language processing (NLP) and natural language generation (NLG) methods to understand and generate dialogues with users. The service should understand each user's different expression in sentence, and should have a conversation that match user's linguistic level.

**Thesis** We contend that there is a better way to provide conversational services for psychiatric counseling. It is better to diagnose the user's mental states, because a one-to-one questionnaire improves accuracy of the psychiatric evaluations. It is also possible to grasp emotional states more accurately based on the multi-modal methods. The service would be useful for emergencies such as suicidal ideation, and long-term mental illnesses such as addiction and bipolar disorder. The conversational service leads to continuous and healthy changes in the patient's self-awareness, personality, and behavior. As a result, the effectiveness of counseling gets better, then the satisfaction to users who need mental health care will improve.

**Plan** We use chat assistant platform to access the counseling service easily. The chat assistant include a chat bot that can applies to various messenger platforms as an interface between conversational service and users. The users who want a psychiatric counselling, can type contents of the consultation as text. In this case, we adapt various natural language processing (NLP) methods to analyze consult contents. In addition, the service can receive various types of additional data from image, video, audio, wearable sensors, and so on. For example, users can send facial pictures, videos, or voice message to the service bot via the messenger. It is also possible to utilize wearable devices that synchronize with the chat bot. The chat bot recognize user current emotions, and convey the results of the emotional recognition to the conversational service. The service put on logs of the results and emotional highs and lows. Through such ongoing dialogues, we monitor continuously and diagnosis the psychiatric problems of users. After that, the service respond appropriately according to user's new inputs using natural language generation (NLG) methods. The service recognizes the users' delicate sentiment changes, and take emergency steps to deal with the urgent cases. At other times, the service suggest useful information to promote mental health, and recommend items for feeling bet-ter. At this time, the conversational service consider clinical psychological and ethical aspects.
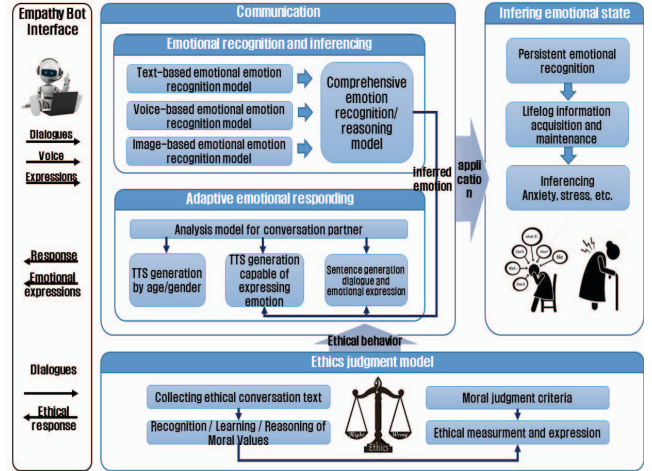


Figure 1.   Overviews of psychiatric counseling service with chatbot.

We sketch a conversational service called psychiatric counseling chatbot for psychiatric counseling with chat assistant in this paper. The service informs the diagnosis and treatment of the psychiatric disorder based on results of emotion recognition obtained from various models. Our vision combines various emotion recognition methods based on multi-modal approach, and response service of psychiatric counseling based on language generation models. Figure 1 shows the overviews of the conversational service for psychiatric counseling.

The paper organize as follows. We survey related techniques to develop the conversational service in following section (Background). After that, we sketch our approaches of the conversational service for psychiatric counseling with partial results of natural language understand, emotional dialogues analysis, and emotional communication in approach section. Finally, we wrap up the paper and presents some on-going and future development in conclusion section..

## II.   BACKGROUND

This paper focus on emotion recognition and monitoring, conversation understanding on chat assistant, Those are the most pivotal technology of the conversational service for psychiatric counseling.

### A.   Emotion Recognition from Various Types of Data

There are various types of data to recognize user's emotion: text [4], image [2], video [2], audio [3], and so on [5].

Reference [4] proposed an unobtrusive emotion recognition approach for affective social communication on mobile devices. They discovered 10 features related to the emotional state of the human user; these features were mainly divided into user behavioral patterns (e.g., typing speed) and the user context (e.g., location). The accuracy of emotional classification showed 67.52% on average for seven emotional states: happiness, surprise, anger, disgust, sadness, fear, and neutral.

Reference [2] presented a framework based on committee machines of deep CNNs and its application to robust FER. They demonstrates on the SFEW2.0 competition data released for the EmotiW2015 challenge. Based on three-level hierarchical committees of total 240 deep CNNs, they achieved 61.6% test accuracy, greatly outperforming the baseline of 39.1 %.

Reference [3] utilized the soft attention mechanism to temporally align the audio and visual streams and fuse these streams in the feature level. In addition, they added the emotion embedding vectors in the output layer of RNN to locate and re-weight the perception attentions in the audiovisual stream.

Reference [5] focused on automatically analyzing a speaker's sentiment in online videos containing movie reviews. They considered adding audio and video features for encoding valuable valence information conveyed by the speaker. Recently, there were many studies on emotion analysis based on multi-modal approaches.

So far, there is a limitation of emotion recognition technology using single information. Currently, most studies use only one type of data such as voice, sound, image data, so those have limitations in performance. Human can infer the other's emotions from information such as voice dialogues, intonation, facial expression, gesture, age, gender, and so on. Furthermore, it has the ability to recognize the degree of emotional stress and respond to the situations. In intelligent assistant, a new model for emotional analysis is needed using such multi-modal information. It is necessary to develop multi-modal method that can obtains considering the strength of each modality, and various information at the same time.

In addition, bi-directional dialogue analysis is not applied. Until now, it is an initial stage in emotion recognition, and there is a limitation of artificial intelligence model, which does not utilize the dialogues. The level of applications has been limited, because the user's emotion and state determines and deduced by one-time recognition from external factors such as speech recognition of words and sentences, and image recognition of faces, peripheral, and location.

### B. Artificial Inteligent for Chat Assistant in Health Care

The chat bot (also called chatterbot) is one kind of intelligent assistant. The chat bot is a computer program, and respond conversation to user via auditory or textual methods. To provide the conversation, the chat bot should understand natural language dialogues and generate natural responses.

Reference [6] developed a user-friendly smart mobile healthcare assistant called MoSHCA to improve patient-doctor interaction and to promote the self-management of chronic diseases by the patients themselves.

They utilized intelligent clinical decision support to require intelligent interpretation of data obtained from the user, and diagnosis or recommending appropriate treatment using probabilistic classification and case-based reasoning. Reference [1] Investigated suitability for a mental health intervention with a simple online Chatbot, as a human surrogate for a health professional. They provided the intervention of specifically alcohol drinking habits assessment of young adults. The trial indicated positive reception of the intervention by users and that availability of chat bot variants with different behavior and sophistication in their conversational ability would further enhance user satisfaction and perceived usefulness.

Emotional intelligence is necessary as an essential function of digital companion. In order for artificial intelligence to function as a digital companion, it must take into emotional aspects similar to personality such as friends and family, beyond rational aspects such as information provision and reasoning. To do this, we need to develop a deep interaction model that recognizes facial expressions continuously during interaction, tracks complex and long-term emotions in various conversations, and expresses appropriate emotions.

Just as humans learn from interaction and communication with others, the emotional intelligence assistant should communicate and learn opinions and emotions with many people. Through this, it is necessary to develop a system that learns common elements firmly and improves oneself by continuously learning the characteristics and emotional state of the individual.

## III. NATURAL LANGUAGE UNDERSTANDING

### A. Understanding Sentence Entailment

It is necessary to understand the sentences in dialogues. Understanding the user's inputs provide a basis to deliver an appropriate response. Each user will have different expressions to use for conversation depending on their linguistic ability and vocabulary level. In the chatbot, method for paraphrasing the various representative expressions applies to understand by the machine using analysis of sentence entailment based on word embedding models.
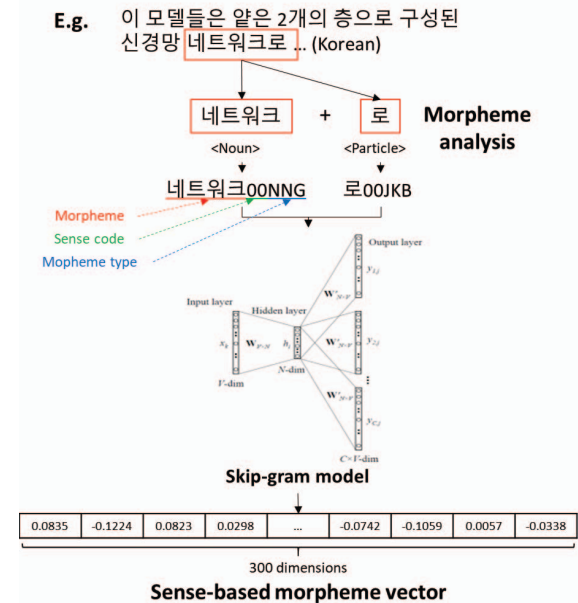


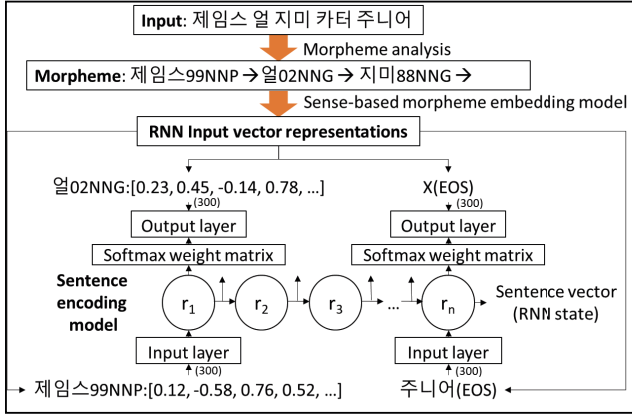Figure 2.   Sense-based morpheme embedding model [7].

Figure 3.    Vector representation using Gated Recurrent Unit (GRU).

First, we collect corpuses of target languages. The word-embedding model learn semantic information of words and represent the word as a vector. The more size of corpuses are collected, the better quality of the word embedding models are generated. In our case, we collected corpuses from Korean Wikipedia, Namuwiki, and news articles.

We learned the models from total 49,864,477 sentences and trained 200 dimensional word vectors of more than 380,000 vocabularies. *Figure* 2 show the embedding model for Korean sentence analysis. Especially, we adopt morpheme embedding which distinguish stems and endings.

Using the word vector, we can collect synonym knowledge of lexical morphemes. The lexical morpheme which have similar semantics have similar word vectors from the synonym knowledge. If the cosine similarity is high among the lexical morphemes, then the morphemes can replace the vocabularies each other. Through a series of process, we can generate paraphrase sentences with decoder and language model trained by above corpuses. The decoder and language model explains in following parts.

### B.  Classification of Sentence Domain and Utterance Intention

In order to the conversational service respond correctly to the user's conversation, the domain analysis and the utterance intention of the sentence entered by the user should be analyzed. Through the GRU [8]-based sentence similarity analysis model, we analyze the domains with high relevance to the input sentences.

Using the model, we can represent the sentences to vectors as *figure* 3. Using the vector can measure distance among domains. It also carries out an analysis of the utterance intention to distinguish whether it is a general conversation or a dialogue that wants to know the correct answer through the utterances classification model that can occur in conversation situations.

As a partial result, we can realized that it is possible to classify sentences effectively according to domains and utterance intents by the vector representation model, as *figure* 4.

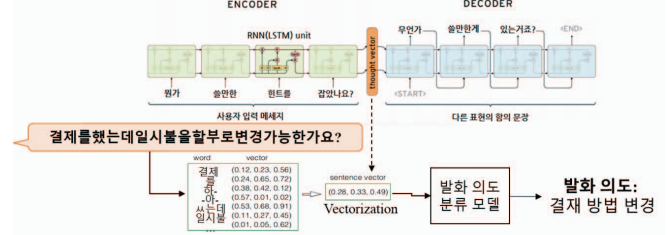

Figure 4.    An Example of sentence domain and utterance intention classification.

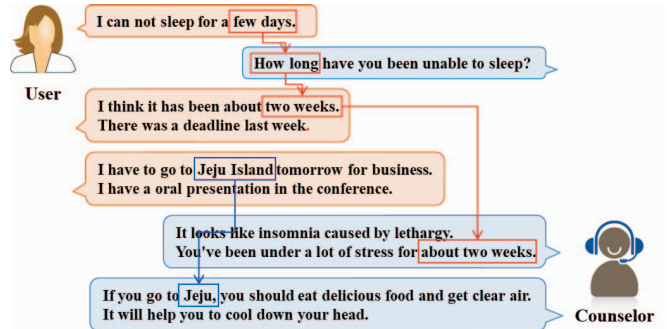### C.  Spatial-Temporal Context Analysis



Figure 5.    An example of spatial-temporal context analysis.

The spatial-temporal context analysis means that analyze explicit and implicit expressions about spatial and temporal context in natural language sentences. The analysis convert unstructured expressions to structured data about the spatial-temporal information. The data is useful to provide a response that matches the spatial-temporal context with the Chatbot. *Figure* 5 shows an example of spatial-temporal context analysis in dialogues. To talk like the example, the Chatbot should remember the temporal and spatial context and respond appropriately according to the context.

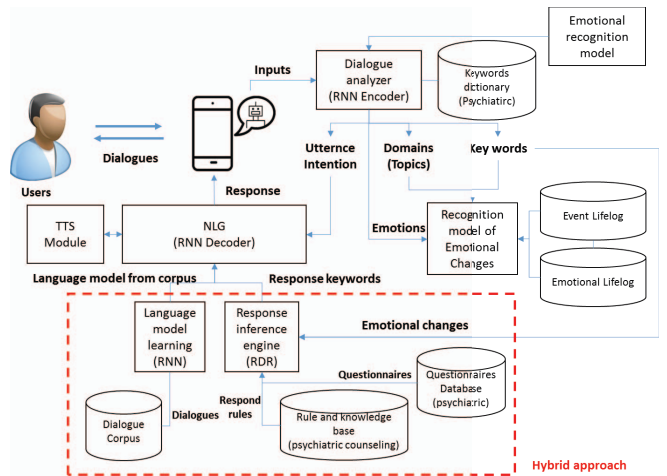## IV.  EMOTIOANL DIALOGUE ANALYSIS
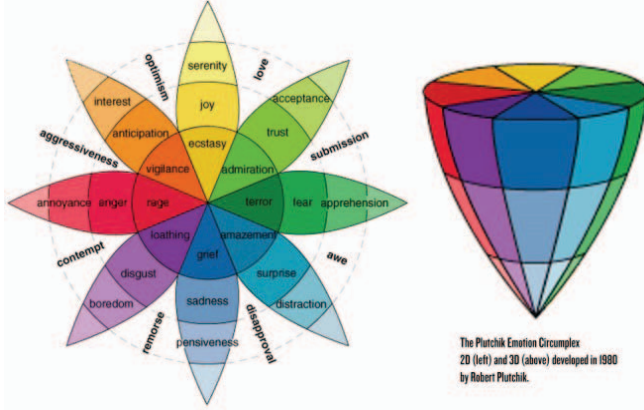


Figure 6.    The counseling chatbot architecture.

Figure 7.   Emotion wheel by Plutchik [9].

The chatbot that provides conversational service for psychiatric counseling collect and summarizes the dialogues of text, voice, and video information to recognize user's emotions. Based on the information, the purpose of the service is to develop a personalized dialog system that communicate emotionally with the user through text, voice, and visual expression. For this, we adopt follows approaches. It is distribution three parts: understanding conversation, emotional recognition, and expression for communications. *Figure* 6 shows the architecture of the chatbot to provide psychiatric counseling.

The conversational service is a type of counseling, so we should understand dialogues between user and Chatbot based on various natural language understanding methods.

### A.   Defining Emotional Expression Model to Categorize

Most of all, we should determine a representative emotional model which can express more than eight kinds of emotions for emotion recognition, learning, and inferencing [9]. The model should be designed to reflect the features of eight types of emotions according to conversation, voice, and image as *figure* 7.

### B.   Collecting Training Data for Emotion Recognition

We should collects training data which contains emotional information from various media such as dramas and radio, and collects also actual dialogues containing emotions through the community and SNS using emotional chat bots to construct corpuses.
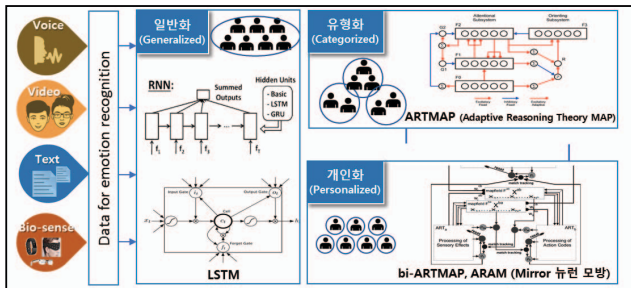


Figure 8.   Fremework for emotional recognition.

*Figure* 8 shows a framework for emotional recognition from various sources: voice, video, text, and bio-sense. Our framework are adopted generalized, categorized, and personalized approaches for the emotional recognition.

### C.   Emotion Recognition and Inference

In this part, a technique uses for judging and inferring the emotions of a user by fusing mutual utterance text, voice, and image information. In addition, additional user information such as facial expression, age, and sex, spatial context, location context, and bio-signals collected via wearable devices. It is possible to expand to various applications by improving emotional intelligence technology based on such the user information.

### D.   Continuous Emotional Mornitoring

Observations of persistent emotional changes are very important in psychiatric counseling to improve the effectiveness of counseling. We apply the user emotion tracking method based on life logging. The maintenance system should construct for continuous information collection. Through this, it is possible to enhance emotional collection based on biological signal sensor and emotion recognition model.

## V.   SENTENCE GENERATION FOR PSYCHIATRIC COUNSELING

It is very important to have a natural conversation in the context of the conversation. If you do not feel that, you are actually talking to someone else in counseling, the effectiveness of counseling will diminish. To solve the problem, we apply a technique to generate natural sentences based on an RNN-based decoder; it generates considering the context qualities of the conversation.

### A.   Emotional Expression and Communications

The method is a user-customized correspondence technology that communicates with users through speech-text-audio-visual representation based on the user's age gender classification and recognized emotions.
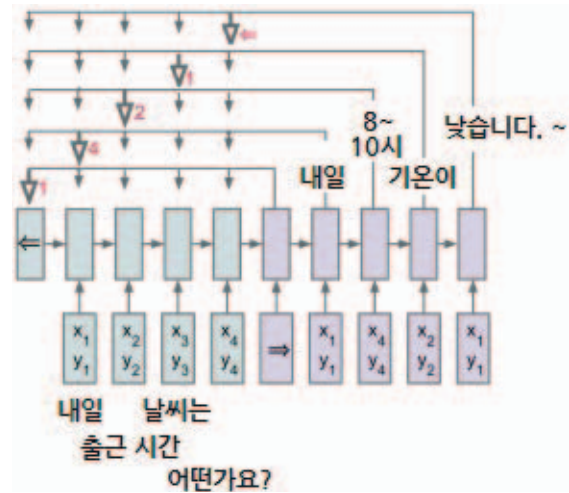


Figure 9.   Pointer network based dialogure response genreation model

The age and gender distinction technique of the user using the image and voice is utilized. The service generate appropriate responses using the information. Accordingly, the corresponding emotional synthesis technique for speech and image can apply according to the user's age and gender. There is also a need for a technique for generating dialogs and expressions according to a user's emotional state.

### B. Psychiatric Intervention

Emotional bots need to apply interactive emotional understanding skills and psychological intervention skills. This can provide sympathetic psychotherapy and treatment services for patients with emotional disorders such as depression, autism, gynecological depression.

Through the distribution and disclosure of emotional intelligence services, it is possible to grasp social anxiety according to age groups and to develop effective countermeasures by grasping the internal state such as general anxiety stress.

### C. Response Generation based on Pointer Network Model

The response generation is a main key of the Chatbot system. Until now, commercial chatbot systems are mostly uses approaches based on rules and templates. When we use these approaches, the chatbot do not answer many cases. If you ask a question using little different expressions, you can get a completely different answer. To solve this problem, we propose a model that generate response sentence based on machine learning approach.

The response generation model learns important keywords in a sentence using Point Network Model [10]. The attention of the sentence learns via feature maps from sentence corpus. The partial result in other Q&A application, the model was very effective to find reply words, as *figure* 9. The *figure* shows that words consisted of answer are pointing each expression in a question. The words in answers are decoded by the question words by machine learning. The model provide substantial response in psychiatric counseling, and we will show the evaluation results in future works.

### D. Ethical Response and Behavior

The criteria for judging human moral values should study together. Psychiatric counseling can affect directly the quality of life of the user [11]. Therefore, when responding to or providing intervention, an optimal action should be taken from an ethical standpoint. If the technology is advanced, it will apply to unethical situations based on the human ethics code model.

## VI. CONCLUSION

In this paper, we present a chatbot for mental healthcare. The chatbot assists psychiatric counseling in dialogues. The service communicates with a user through dialogues and conducts psychiatric counseling.

To understand the dialogues and recognize user's emotion, the service apply various emotional intelligence techniques: a multi-modal emotion recognition from conversation content, intonation, and facial expression, an intelligent corresponding such as psychiatric case-based reasoning and long-term monitoring, and ethics judgment, etc. The techniques enable continuous observation of user's emotional changes sensitively. As a result, the effectiveness of counseling gets better, then the satisfaction to users who need mental health care will improve.

In further study, we will apply the conversational service to game addiction cases of teenage, and investigate clinical results and user satisfaction.

### REFERENCES

[1] D. Elmasri, and A. Maeder, "A Conversational Agent for an Online Mental Health Intervention," In proc. of International Conference on Brain and Health Informatics, pp. 243-251, 2016.

[2] B. K. Kim, J. Roh, S. Y. Dong, and S. Y. Lee, "Hierarchical committee of deep convolutional neural networks for robust facial expression recognition," Journal on Multimodal User Interfaces, pp. 1-17, 2016.

[3] L. Chao, J. Tao, M. Yang, Y. Li, and Z. Wen, "Audio Visual Emotion Recognition with Temporal Alignment and Perception Attention," arXiv preprint arXiv:1603.08321, 2016.

[4] H. Lee, Y. S. Choi, S. Lee, and I. P. Park, "Towards unobtrusive emotion recognition for affective social communication," In proc. of 2012 IEEE Consumer Communications and Networking Conference, pp. 260-264, 2012.

[5] M. W?llmer, F. Weninger, T. Knaup, B. Schuller, C. Sun, K. Sagae, and L. P. Morency, "Youtube movie reviews: Sentiment analysis in an audio-visual context," IEEE Intelligent Systems No. 28 Vol. 3, pp. 46-53, 2013.

[6] A. Hommersom, P. J. Lucas, M. Velikova, G. Dal, J. Bastos, J. Rodriguez, M. Germs, and H. Schwietert, "Moshca-my mobile and smart health care assistant," In proc. of e-Health Networking, Applications & Services (Healthcom), pp. 188-192, 2013.

[7] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.

[8] K. Cho, B. Van Merri?nboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," arXiv preprint arXiv:1406.1078, 2014.

[9] R. Plutchik, "Emotions and Life: Perspectives from Psychology," Biology, and Evolution, Washington, DC: American Psychological Association, 2002.

[10] T. H. Wen, D. Vandyke, N. Mrksic, M. Gasic, L. M. Rojas-Barahona, P. H. Su, and S. Young, "A network-based end-to-end trainable task-oriented dialogue system," arXiv preprint arXiv:1604.04562, 2016.

[11] L. W. Roberts, J. Battaglia, and R. S. Epstein, "Frontier ethics: Mental health care needs and ethical dilemmas in rural communities," Psychiatric Services, 1999.

.