The ChatBot Feels You – A Counseling Service Using Emotional Response Generation

Dongkeon Lee, Kyo-Joong Oh, Ho-Jin Choi
School of Computing, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea
Email: {hagg30, aomaru, hojinc}@kaist.ac.kr

***Abstract*—Early study tries to use chatbot for counseling services. They changed drinking habit of who being consulted by leading them via intervene chatbot. However, the application did not concerned about psychiatric status through continuous conversation with user monitoring. Furthermore, they had no ethical judgment method that about the intervention of the chatbot. We argue that more reasonable and continuous emotion recognition will make better mental healthcare experiment. It will be more proper clinical psychiatric consolation in ethical view as well. This paper suggests a introduce a novel chatbot system for psychiatric counseling service. Our system understands content of conversation based on recent natural language processing (NLP) methods with emotion recognition. It senses emotional flow through the continuous observation of conversation. Also, we generate personalized counseling response from user input, to do this, we use additional constrains to generation model for the proper response generation which can detect conversational context, user emotion and expected reaction.**

Keywords— conversational service; response generation; deep learning;

I. INTRODUCTION

Emotional recognition of human has been a long research topic. Recently many studies show artificial intelligence (AI) methods are adequate approach. To build various emotion classification models, a number of emotional-labeled data are used in the studies. For instance, in static image processing convolution neural network [1], in temporal time domain, recurrent neural network [2], especially for the machine translation, attention network [2] [3] are known. Along with the technical advance, the training data are also differentiation to image [1], video [1] [2], audio [3] and text [4]. In addition, some studies [2] [4] showed hybrid approached multi-modal classification. The studies improved result for emotion recognition significantly.

However, not many applications applied the brand-new emotion recognition techniques. Nowadays, for example, Apple Siri, Google now, Samsung S-Voice. These are some known intelligent assistant services. The basic idea of these services are they respond to the users' inputs, such as queries of voice or text, and they recommend useful information to the users. But, the services just apply very simple natural language processing (NLP) techniques. The key applications are not very varied yet. So, the healthcare application should be studied together with the intelligent assistant [5]. Some chatbot and chat assistant are bright promising in the market.

In this paper, we introduce a novel chatbot application which provide mental healthcare counseling service based on above natural language processing and emotion recognition methods in chat assistant platform which consist of the user sensitive emotion and context extraction.

II. RELATED WORK

This paper focus chat assistant which can recognize and monitoring the human emotion and understand the natural language conversation, the most crucial technologies in the conversational psychatric counseling service.

A. Emotion Recognition

There In previous work, there are several types recognition about user's emotion: text[3], image and video[1] and audio[2] [3] proposed an emotion recognition approach for mobile social network services. They found 10 features that indicates the emotional state of the user; those features were mostly determined by the behavioral user patterns and the contextual user pattern (e.g., each typing speed and location). Emotional classification showed 67.52% accuracy on average for the 7-emotional states that follows: happiness, surprise, anger, disgust, sadness, fear, and neutral. [1] showed committee machines a framework that has sturucture of deep CNNs. It show robust face expression recognition. The paper demonstrated those model on the SFEW2.0 competition dataset that released for the EmotiW2015 challenge. The model structures are 3 levels of hierarchical committees. They achieved test accuracy of 61.6 %, which highly outperforms the baseline of 39.1 %. In [2], they align temporal the audio and visual streams by utilizing the soft attention mechanism. They added emotion embedding vectors in output layer of RNN. It locates and re-weight the perception attentions along the audio visual stream.

There are limitations in performance for previous emotion recognition methods using a single feature only. Human can inference emotions of the others from integrated information. Also, human recognizes the degree of emotional state and responses from the circumstance condition. So, for the intelligent assistant, integrating multi-modal information is inevitable.

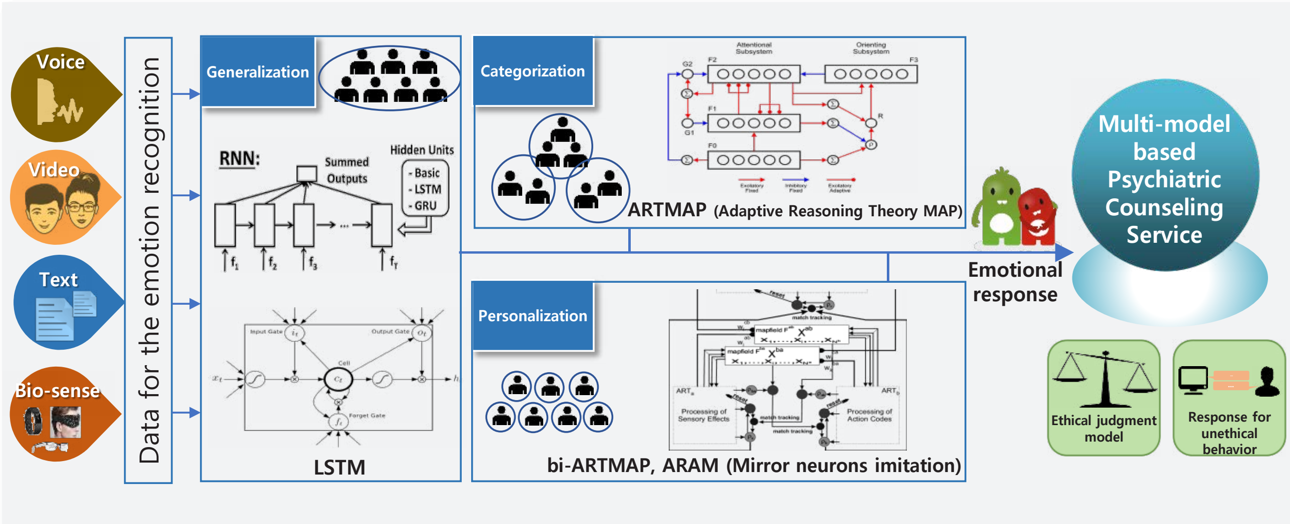


Fig. 1. Framework for emotional recognition.

The integration can be obtained by adjusting the strength of each model from the various informations in multiple data source at the same time. Also, they are used just one-directional dialogue analysis. They only focused on the emotional analysis itself. The root of the limiation is, until now, user emotions are inferenced by only one-time recognition from external factors.

B. *Chat Assistant for Mental Healthcare*

Chatbots broadly are used for the intelligent assistant applications. In common, they generate responses from the user's input. The chatbot need to have a capacity to analyze natural language dialogue. [5] introduces a smart mobile healthcare assistant. They improve patient-doctor interactions. [6] argues that the chatbot can substitute professional counselor by intervening alcohol drinking habits.

Emotional intelligence is necessary as an essential function of digital companion. To do this, we need to develop a deep interaction model that recognizes complex and long-term emotions in various conversations continuously. Just as human counselor need to learn from many interaction and communication to react counselee properly, the emotional intelligent 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.

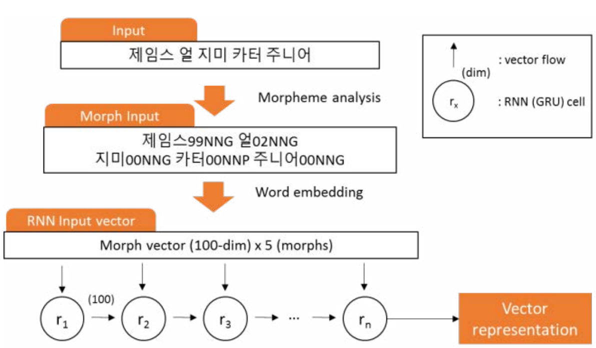


Fig. 2. Vector representation using gated recurrent unit (GRU).

III. Emotional Response Generation

The response generation that we introduced for providing conversational services for mental healthcare collects and analyzes, integrates the input dialogues consist of texts to recognize the emotion of user. Our purpose is developing a user sensitive dialog system that can comunnicate user over time with continuous contextual self-awareness through those information. For this, we seperate this into four steps: feature extraction, response decision, generating sympathetic response, generating informative response.

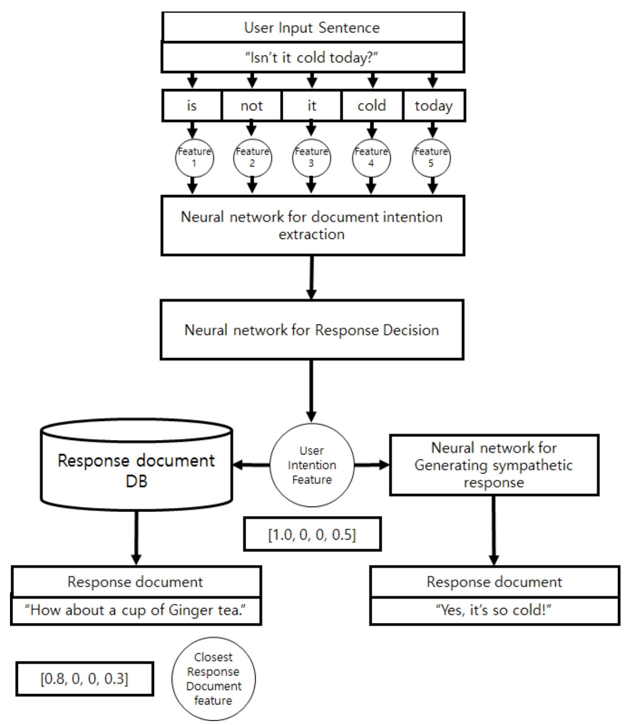


Fig. 3. Framework for response generation.

Figure 3 shows how the response generation framework works. The upper level of the model is the user intention extraction. It can be replaced by previous steps to automate this learning procedure. The bottom level is response generation. Response document are store in the database that consist of KB and tagged web document sets to find proper pair form user intention, the document has their target intention. The reason why we choose the database to store response document is the probabilistic model that we used for the user emotion recognition or prediction is quite noisy to generate precise recommendation for the user intention. So, on our system, we combine probabilistic model to recognize or predict the user emotion and provide the detail reaction by the structured or semi-structured data storage.

A. *Emotional Recognition Based on NLP*

For the conversational mental healthcare service, initially, the system should understand the natural language sentence. Since every user had different linguistic ability. We need to normalize data into the fixed length vector representation. Using GRU-based sentence analysis [7], the model extract the utterance intention to the representation. The similarity anlysis can be done with cosine simularity measuring. Our model shown in Figure 2, which can measure distance among the sentences along the domain information.

The method that we used for recognize and inference the user emotions from fusing mutual utterance information such as text. Additionally, user information such as location, sex, age, facial expression or any other personally generated signals are can be collected through the wearable devices for more accurate inference.

We should choose 8-kind of emotions from a representative emotional which originally more than 8 emotions to simlify the problem [8]. We collect the data form the public media which drama and radio scripts and SNS dialogues which contains manually descripted emotions. Figure 1 shows the framework for the emotional recognition from various sample inputs.

The system also track the user emotion based on life log system. The maintenance system collects the bio information continously to track the user context. It enhances emotional collection which used to emotional recognition model.

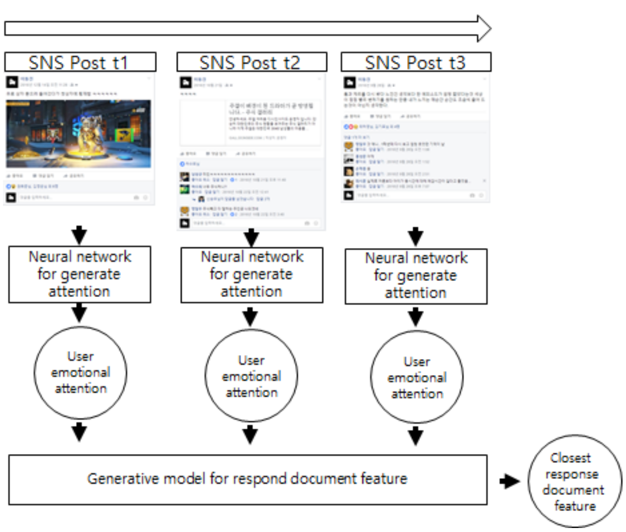


Fig. 4. Framework for Emotional feature Extraction.

We also focused about the generation of emotional text. Since chatbot offers psychiatric counselling, the bot need to express emotional state properly not just simply recognize it. The generation of emotional text is much different problem compares to the recognition of the emotion. The model understand not just the conversion of input features to high level features, it need to compress the high level features to generate emotional response properly in limited memory. So, we not just use the generated intention features from the model in Figure 3 to inference closest response document. We try to pre-recognize the emotional state from the 6 emotional state that we have chosen for generate document features. [2] said the bottleneck problem is severe problems in long length of input, which can be thought as long term time dependent features in our model. We expect it frees the model from the generation bottleneck. Its intuition came from attention mechanism in [2], which showed great performance in the machine translation. Also, it is possible to integrate those extra features to the previous model easily because the interface of the previous model is simple, it just need the feature vector to find the closest document.

Figure 4 shows that model which generate emotional features from the model should collect the data from the SNS which contains various information in it. It relative to the user emotion via the social network, for example, text, picture, and social activities. We use transformed data in previous steps and put this to our emotional attention based model to generate contextual sensitive generation for proper response.

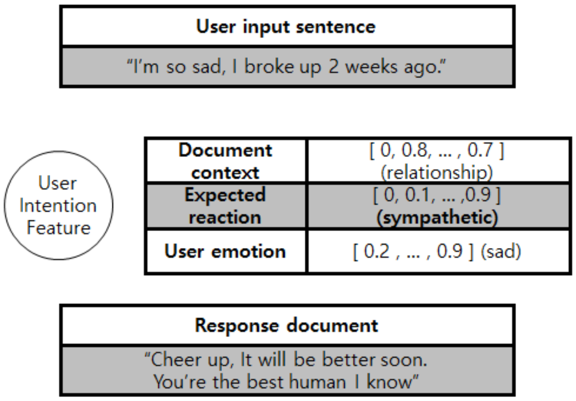
In addition, the study of the criteria for the human moral judgement on this system is needed. The psychiatric counseling can affect to the user's life directly. Before the model generates the respond reation to providing intervention, the respond must be carefully taken from an ethical standpoint. Aside the technology advance, the system can face the unethical situations through the learning process in ethical aspect of the human. For these ethical achievement, even form the data collection, to the response generation and feedback, the whole process should be done by continous attention with ethical view.

B. *Personalized Response Generation*

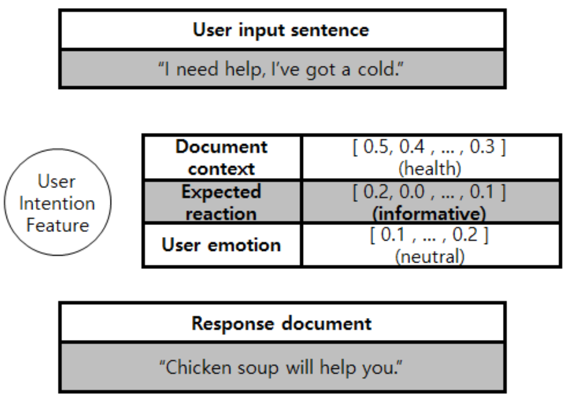
The user's personal information like age and gender improves personalized response generation method for classification and recognition of emotions in probabilistic sense. The service can generate appropriate response using the user input features. In the same time, the user's emotion extraction be applied according to the user's personal information like age and gender.

Also, for the proper response generation, system need to collect pair dataset of user intention and proper response we expect, the intention previous steps and the right response can be obtained from both manually and automatically which came from the structural data of target domain knowledge pair, such as, knowledge base (KB) or Web document.

The user intention consists of two component briefly, one is the document context and another is user emotion. The possible issue of this system for the emotion extraction and the document generation is that how to deal the user intention and the document context features together. In the model above can face the situation that only the user intention or the document feature affects the model output. If one of those feature dominates the feature affection, the effectiveness of the model drops severely.



(A) Sympathetic response



(B) Informative response

Fig. 5. Detailed feature selection mechanism. In (A), the response generated from probabilistic model that deciced the input sentence expect sympathetic reaction. In costrast, (B), expect informative reaction.

So, we planned detailed feature selection model, we put constraint to the user intention feature by putting expected reaction feature. It helps recognize the user intention whether he or she wants emotional sympathetic response, for instance, sympathize, sorry, be astonished, which generated from probabilistic model or more informative response that gives user precise information from the (semi-)structured data storage as we explained above.

Figure 5 said how those feature selection works, the feature selection done automatically from the one of labeled user intention that indicates what user wants from the machine, in view of sympathetic or informative.

IV. Conclusion

We introduced an novel chatbot application which provide conversational mental healthcare service based on emotion recognition methods and chat assistant platform which consist of the context sensitive advanced natural language-based technique to provide personalized response generation continously. To understand the natural language input sentences and recognize user's emotions, system got 4 levels of hierarchy. The system enables the natural language processing to understand user state and observation of continuous user's emotional changes sensitively. Consequently, the experience of users who need mental healthcare will be improved by our approaches. In future works, we apply the emotional generation to natural language reponse sentence not just feature extraction.

Acknowledgment

This work was supported by the Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. R0101-16-0062, Development of Knowledge Evolutionary WiseQA Platform Technology for Human Knowledge Augmented Services).

References

[1] 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.

[2] 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.

[3] 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.

[4] M. Wöllmer, 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 **28**(3), pp. 46-53, 2013.

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

[6] D. Elmasri, 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.

[7] K. Cho, B. Van Merriënboer, 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.

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