INTERSPEECH 2017: Show & Tell Contribution
August 20–24, 2017, Stockholm, Sweden

Nora the Empathetic Psychologist

Genta Indra Winata[1], *Onno Kampman*[1], *Yang Yang*[1], *Anik Dey*[2], *Pascale Fung*[1,2]

[1]Human Language Technology Center
Department of Electronic and Computer Engineering
Hong Kong University of Science and Technology, Hong Kong
[2]EMOS Technologies Inc.

[giwinata, opkampman, yyangag]@connect.ust.hk, pascale@ece.ust.hk

Abstract

Nora is a new dialog system that mimics a conversation with a psychologist by screening for stress, anxiety, and depression. She understands, empathizes, and adapts to users using emotional intelligence modules trained via statistical modelling such as Convolutional Neural Networks. These modules also enable her to personalize the content of each conversation.

Index Terms: virtual psychologist, dialog system, human-agent interaction, affective computing

1. Introduction

Here we present **Nora**, a new dialog system with a web-based virtual agent (VA) in the role of a psychologist. Mental health issues are very common in the world today. Yet, seeking professional psychological help is still stigmatized in many cultures. We explore the potential of our platform to motivate and assist people. We emphasize that our platform is not intended to replace human therapists, but rather serves as an enhancement to in-person sessions.

Nora is capable of recognizing stress, emotions, personality, and sentiment from speech. She learns from every user conversation through machine learning algorithms. In particular, an emotional intelligence (EI) module is incorporated to enable emotion understanding, empathy, and adaptation to users. The design follows the abilities model of EI theory developed in the 1990s [1], where appraisal and understanding emotions lay the foundation for more sophisticated abilities like using and managing emotions. It has been shown that VAs portraying empathy can increase user engagement and satisfaction [2], [3]. Therefore, we adjust the conversation flow to these parameters and create a personalized experience. Nora can also screen for depression, anxiety, and stress based on the DASS-21 model.

2. System description

2.1. Design

Nora is an empathetic dialog system that takes audio and facial image of the user as input. The system records the user's speech using built-in microphone and captures facial image from a web camera. It uses existing libraries for Automatic Speech Recognition and Text-to-Speech. Empathetic modules are used to evaluate stress level, sentiment, emotion, and personality. A mixed-initiative dialog system is designed with dialog manager to personalize actions to each user and remember the input and states. The language generation component generates text response corresponds to the selected action.

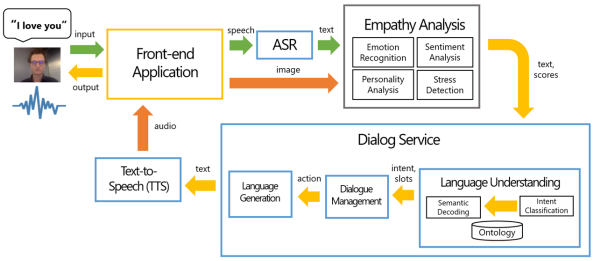

Figure 1: *Schematic overview of the dialog system flow.*

2.2. System architecture

We built the system using microservices architecture to ensure the modularity and scalability in building and running each module. The client-side application will send the user input to the server that handles the request, by calling multiple services such as dialog service, empathy analysis, and aggregating the results (see Figure 1). Audio input is analyzed by the dialog service to get the user's intent. It also determines the response based on four empathetic scores: stress level, emotion, perceived personality, and sentiment scores. The dialog service consists of natural language understanding (NLU), dialog management, and language generation modules. The NLU module identifies user intent and slots. For each dialog turn, intent and slots are evaluated by the dialog management module. The module remembers all past dialog turns, and evaluates the intent and slots to get the desired action. The language generation module translates actions into text responses.

2.3. Emotional intelligence modules

The core technologies behind Nora's emotional intelligence are our four empathy modules. These enable perception and understanding of social cues. All modules are developed in-house.

2.3.1. Stress detection from audio

We built a model that can detect stress from spoken utterances. For this we use the Natural Stress Emotion corpus [4]. The corpus contains spoken utterances in English, Mandarin, and Cantonese, of students answering the same set of 12 questions. The utterances are either labeled as "stressed" or "unstressed".

Previously, a stress detection classifier was constructed using an SVM [5]. This method requires audio feature extraction and is therefore not suitable for real-time systems like ours. We created a new model using a Convolutional Neural Network (CNN), which automatically extracts relevant representations of the raw audio input. This deep learning architecture is shown in

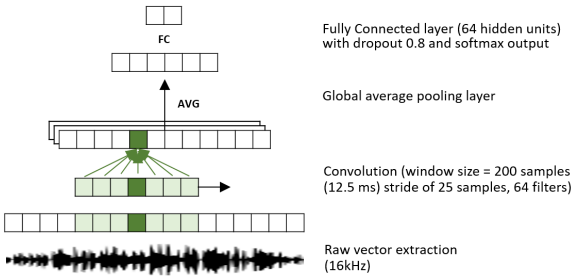

Figure 2: *Convolutional Neural Network architecture.*

Figure 2.

It takes as input a raw audio sample, with a sampling rate of 16 kHz. The first convolutional layer is applied directly to the audio sample $\mathbf{x}$:

$$\mathbf{x}_i^{\mathrm{C}} = \mathrm{ReLU}(\mathbf{W}_{\mathrm{C}}\mathbf{x}_{[i,i+v]} + \mathbf{b}_{\mathrm{C}}) \quad (1)$$

where v is the convolution window size. We apply a window size of 200, sampling rate corresponds to 12.5 ms, and output depth of 25, and move the convolution window with a step of 25. The convolutional layer is followed by a global average-pooling layer. This layer's output is fed to a fully-connected layer with dropout 0.8, and a final softmax layer that serve as a classifier.

2.3.2. Automatic emotion detection from audio

For emotion recognition from speech, we deploy a similar model as used for stress detection. The model is trained using a CNN on raw audio input labeled for six different emotions. It outputs a continuous predicted score for each of the emotions. These are used to give a more empathetic response.

2.3.3. Counseling functionality

After Nora detects elevated levels of stress, anxiety, or depression, she will offer relieve through counseling.

3. User experience and interface

Nora has the emotional intelligence and empathetic capabilities needed in a virtual psychologist. We designed the web interface, dialog, and animations based on insights from human-agent interaction research. Simple expressions of emotion via text, speech, or face have shown to already be effective [3].

3.1. Web interface design

The web interface is optimized for understandability and usability. Users can log into our system via online accounts, after which they reach the dialog session page (shown in Figure 3). We added an avatar for Nora and display the conversation in the style of a messaging app to increase user engagement. The empathy modules' output is visualized throughout the conversation.

3.2. Dialog design

In each conversation, Nora will ask a series of questions. The first questions are light-hearted, so users feel relaxed [2]. They then get progressively more intimate and exposing to invoke emotions. The system sets priorities to detect suicidal thoughts

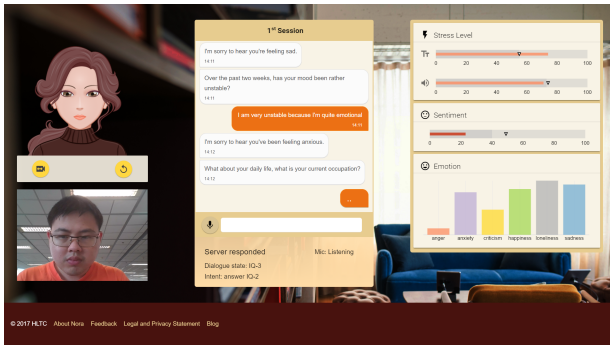

Figure 3: *Dialog session page layout.*

and abusive language. In face of emotion-loaded and affect-sensitive information, Nora responds empathetically prior to proceeding to the next question (see sample dialog below). To further engage users, Nora animates her facial expressions based on detected emotions. Throughout each session she screens for stress-related, anxious, and depressive expressions, and summarizes her insights after each session.

3.2.1. Example scenario

Nora: "How is your work going recently? Any deadlines?"
User: "I am worried about paper deadlines."
...
Nora: "I sense that you are quite stressed. How do you usually manage your stress?
User: "I eat a lot of chocolate. But after I feel guilty."
Nora: "I can empathize with that."

4. Conclusions

Nora the Virtual Psychologist screens for stress, anxiety, and depression during conversations using emotional intelligence modules. She is aware of each user's stress level and mood and uses this ability to respond empathetically. We will open up the system to the public soon and explore applications for the mental health community.

5. References

[1] P. Salovey and J. D. Mayer, "Emotional intelligence," *Imagination, cognition and personality*, vol. 9, no. 3, pp. 185–211, 1990.

[2] P. Fung, D. Bertero, Y. Wan, A. Dey, R. H. Y. Chan, F. B. Siddique, Y. Yang, C.-S. Wu, and R. Lin, "Towards empathetic human-robot interactions," *arXiv preprint arXiv:1605.04072*, 2016.

[3] Y. Yang, X. Ma, and P. Fung, "Perceived emotional intelligence in virtual agents," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2017, pp. 2255–2262.

[4] X. Zuo, L. Lin, and P. Fung, "A multilingual database of natural stress emotion," in *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC12)*, 2012, pp. 1174–1178.

[5] X. Zuo and P. N. Fung, "A cross gender and cross lingual study on acoustic features for stress recognition in speech," in *Proceedings 17th International Congress of Phonetic Sciences (ICPhS XVII), Hong Kong*, 2011.