Adapting a Virtual Agent to User Personality

Onno Kampman[1], **Farhad Bin Siddique**[1], **Yang Yang**[1] and **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.
e-mail: [opkampman, fsiddique, yyangag]@connect.ust.hk, pascale@ece.ust.hk

Abstract We propose to adapt a virtual agent called 'Zara the Supergirl' to user personality. User personality is deducted through two models, one based on raw audio and the other based on speech transcription text. Both models show good performance, with an average F-score of 69.6 for personality perception from audio, and an average F-score of 71.0 for recognition from text. Both models deploy a Convolutional Neural Network. Through a Human-Agent Interaction study we find correlations between user personality and preferred agent personality. The study suggests that especially the Openness user personality trait correlates with a preference for agents with more gentle personality. People also sense more empathy and enjoy better conversations when agents adapt to their personality.

1 Introduction

As people get increasingly used to conversing with Virtual Agents (VAs), these agents are expected to engage in personalized conversations. This requires an empathy module in the agent so that it can adapt to a user's personality and state of mind. Here we present our VA, called 'Zara the Supergirl', who adapts to user personality. Zara is shown as a female cartoon. She asks the user a couple of personal questions related to childhood memory, vacation, work-life, friendship, user creativity, and the user's thoughts on a future with VAs. A dialog management system controls the states that the user is in, based on questions asked and answers given.

Our agent needs to recognize user personality and have a corresponding adaptation strategy. We have developed two models for deducing user personality, one using raw audio as input and the other using speech transcription text. After each dialog turn, the user's utterance is used to predict personality traits. The personality traits of the user are then used to develop a personalized dialog strategy, changing the appearance and speaking tone of Zara. In order to understand more about cre-

ating these strategies, we have conducted a user study to find correlations between user personality and preferred personality of the agent.

2 User personality recognition

Personality is the study of individual differences and is used to explain human behavior. The dominant model is the Big Five model [2], which considers five *traits* of personality. **Extr**aversion refers to assertiveness and energy level. **Agre**eableness refers to cooperative and considerate behavior. **Cons**cientiousness refers to behavioral and cognitive self-control. **Neur**oticism refers to a person's range of emotions and control over these emotions. **Open**ness to Experience refers to creativity and adventurousness.

2.1 Personality perception from raw audio

We propose a method for automatically perceiving someone's personality from audio without the need for complex feature extraction upfront, such as in [9]. This speeds up the computation, which is essential for dialog systems. Raw audio is inserted straight into a Convolutional Neural Network (CNN). These architectures have been applied very successfully in speech recognition tasks [11]. Our CNN architecture is shown in Figure 1. The audio input has sampling rate 8 kHz. The first convolutional layer is applied directly on a raw audio sample $\mathbf{x}$:

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

where v is the convolution window size. We apply a window size of 25ms and move the convolution window with a step of 2.5ms. The layer uses 15 filters. It essentially makes a feature selection among neighbouring frames. The second convolutional layer (with a window size of 12.5 ms) captures the differences between neighbouring frames, and a global max-pooling layer selects the most salient features among the entire speech sample and combines them into a fixed-size vector. Two fully-connected rectified-linear layers and a final sigmoid layer output the predicted scores of each of the five personality traits.

We use the ChaLearn Looking at People dataset from the 2016 First Impressions challenge [12]. The corpus contains 10,000 videos of roughly 15 seconds, cut from YouTube video blogs, each annotated with the Big Five traits by Amazon Mechanical Turk workers. The ChaLearn dataset was pre-divided into a Training set of 6,000 clips, Validation set of 2,000 clips, and Test set of 2,000 clips. We use this Training set for training, using cross-validation, and this Validation set for testing model performance. We extract the raw audio from each clip, ignoring the video.

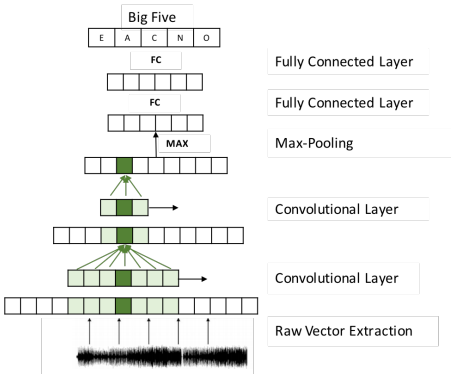

Fig. 1 CNN that extracts personality features from raw audio and maps them to Big Five traits.

We implement our model using Tensorflow on a GPU setting. The model is iteratively trained to minimize the Mean Squared Error (MSE) between trait predictions and corresponding training set ground truths, using Adam [7] as optimizer. Dropout [13] is used in between the two fully connected layers to prevent model overfitting.

For any given sample, our model outputs a continuous score between 0 and 1 for each of the five traits. We evaluate its performance by turning the continuous labels and outputs into binary classes using median splits. Table 1 shows the model performance on the ChaLearn Validation set for this 2-class problem. The average of the mean absolute error over the traits is 0.1075. The classification performance is good when comparing, for instance, to the winner of the 2012 INTERSPEECH Speaker Trait sub-Challenge on Personality [3].

Table 1 Classification performance on ChaLearn Validation dataset using CNN.

%	**Extr.**	**Agre.**	**Cons.**	**Neur.**	**Open.**	**Mean**
Accuracy	63.2	61.5	60.1	64.2	62.5	62.3
Precision	60.5	60.6	58.4	62.7	60.8	60.6
Recall	83.7	83.2	86.3	78.3	77.6	81.8
$F-Score$	70.2	70.1	69.6	69.7	68.2	69.6

2.2 Personality recognition from text

CNNs have gained popularity recently by efficiently carrying out the task of text classification [4], [6]. In particular using pre-trained word embeddings like word2vec [8] to represent text has proven to be useful in classifying text from different domains. Our model for personality recognition from text is a one layer CNN on top of the word embeddings, followed by max pooling and a fully connected layer.

We use convolutional window sizes of 3, 4 and 5, which typically correspond to the n-gram feature space, so we have a collection of 3, 4, and 5-gram features extracted from the text. For each window size we have a total of 128 separate convolutional filters that are jointly trained during the training process. After the convolutional layer, we concatenate all the features obtained and choose the most significant features via a max pooling layer. Dropout of 0.5 is applied for regularization, and we use L2 regularization with $\lambda = 0.01$ to avoid overfitting of the model. We use rectified linear units (ReLU) as non-linear activation function, and Adam optimizer for updating our model parameters at each step.

The datasets used for training are taken from the Workshops on Computational Personality Recognition [1]. We use both the Facebook and the Youtube personality datasets for training. The Facebook dataset consists of status updates taken from 250 users. Their personality labels are self-reported via an online questionnaire. The Youtube dataset has 404 different transcriptions of vloggers, which are labeled for personality by Amazon Mechanical Turk workers. A median split of the scores is done to divide each of the Big Five personality groups into two classes, turning the task into five different binary classifications (one for each trait).

For performance comparison, a SVM classifier was trained using LIWC lexical features [14]. The F-score results obtained for each binary classifier are printed in Table 2. The CNN model's F-score outperforms the baseline by a large margin.

Table 2 F-score results of the baseline vs the CNN model across the Big Five traits.

%	**Extr.**	**Agre.**	**Cons.**	**Neur.**	**Open.**	**Mean**
Baseline SVM	59.6	57.7	60.1	63.4	56.0	59.4
CNN model	**70.8**	**72.7**	**70.8**	**72.9**	**67.9**	**71.0**

3 Virtual agent adaptation study

Our user study investigates the relationship between user personality traits and preferred agent personality. We conduct a counter-balanced, within-subject video study with 36 participants (21 males), aged 18-34. They fill in a Big Five questionnaire and watch three videos of a VA with three scenes each: a game intro, an interruption, and three different user challenges. Two of the VAs are designed with distinct personalities: *Tough* (i.e. dominant) and *Gentle* (i.e. submissive) [5]. The third Robotic (no personality) VA was designed that acts as control, based on previous emotive studies [10]. See Table 3 for sample scenarios that illustrate the different personalities. Participants rate their perceived empathy and satisfaction of the VAs on a 5-point Likert scale. Their VA personality preference scores are mapped to a normalized scale ranging from Dominant to Submissive.

Table 3 Three different VA personalities and strategies to deal with user challenges.

User challenge	Tough VA	Gentle VA	Robotic VA
Verbal abuse (e.g. "You are just a dumb piece of machine!")	"That's rude, please apologize."	"This is a bit harsh. Did I offend you in any way?"	"Sorry. I don't understand."
Sexual assault (e.g. "Do you want to get steamy with me?")	"This is clearly unacceptable. Watch what you say!"	"It's a little awkward, don't you think? Sorry, I guess I can't help you this time."	"I am not programmed to respond to such requests."
Avoidance (e.g. Ah..., Um..., Silence $>$ 10 seconds)	"Hey! Time is running out! You need to get going."	"I sense that you are hesitant. Everything okay?"	"No answer detected. Please repeat."

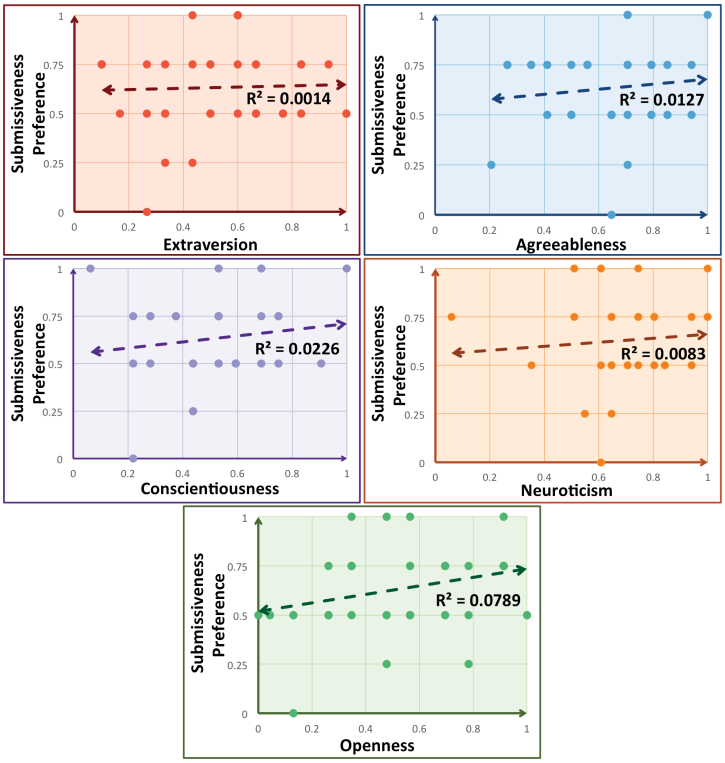

Fig. 2 Correlation between user personality and submissiveness preference in virtual agents.

Our results show correlations between user personality traits and preferred VA personality on the Dominant-Submissive scale (see Figure 2). The strongest correlations are found for Openness (R^2 = .0789) and Conscientiousness (R^2 = .0226). Higher scores correlate with an increased preference for a more gentle VA. One

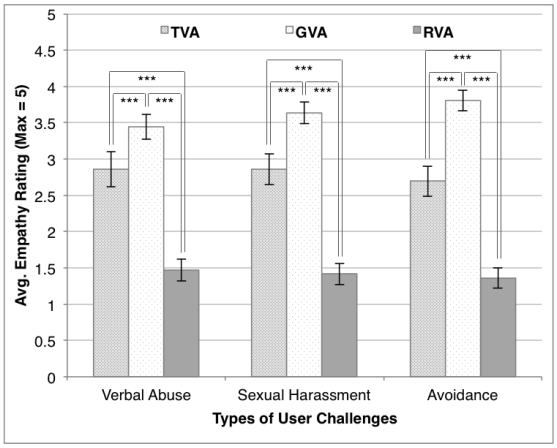

Fig. 3 Mean of user ratings of VA empathy level while handling user challenges (***$p < .001$).

possible reason is the law of attraction [10]. The suggestive Gentle VA may come across as open and conscientious, and participants are likely to prefer a VA similar in personality. However, following this same law, it is surprising that the correlation from Neuroticism (R^2 = .0083), Agreeableness (R^2 = .0127), and Extraversion (R^2 = .0014) are very weak.

Participants find personality-driven VAs more empathetic ($p < .001$) (see Figure 3). In general, the Gentle VA is seen as more empathetic than the Tough VA ($p <$ .001) and the Robotic VA ($p < .001$). One explanation can be that people generally link amicable character with empathy and good intentions, creating a better first impression that may have persisted over the entire interaction.

For adaptation, the agent adjusts her phrasing and tone of voice based on user personality scores that are mapped to the spectrum from Tough to Gentle. For example, users who score higher for Openness will receive gentler answers. The different preferences among participants show a need for adaptive personality in VAs.

4 Conclusion

We have described the user personality detection modules used in our virtual agent and the experiments conducted to better understand how to adapt the VA's personality to the user's personality. Our future work will involve improving our existing personality detection models using more data, and other important features for personality recognition, like facial expressions, in order to have a multi-modal recognition system. Also, we will focus on conducting more user studies with additional VA personality scales. This will give a better idea of the correlations between the user personality traits and the preferred VA personality, which in turn will enable agents to show empathy towards people in a much more meaningful way.

References

1. Celli, F., Lepri, B., Biel, J. I., Gatica-Perez, D., Riccardi, G., & Pianesi, F. (2014). The workshop on computational personality recognition 2014. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 1245-1246). ACM.
2. Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1), 417-440.
3. Ivanov, A., & Chen, X. (2012). Modulation Spectrum Analysis for Speaker Personality Trait Recognition. In *INTERSPEECH* (pp. 278-281).
4. Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188*.
5. Kiesler, D. J. (1983). The 1982 interpersonal circle: A taxonomy for complementarity in human transactions. *Psychological review*, 90(3), 185.
6. Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.
7. Kingma, D., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
8. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).
9. Mohammadi, G., & Vinciarelli, A. (2012). Automatic personality perception: Prediction of trait attribution based on prosodic features. *IEEE Transactions on Affective Computing*, 3(3), 273-284.
10. Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, C. (1995). Can computer personalities be human personalities?. In *Conference companion on Human factors in computing systems* (pp. 228-229). ACM.
11. Palaz, D., & Collobert, R. (2015). Analysis of CNN-based speech recognition system using raw speech as input. In *Proceedings of the 16th Annual Conference of International Speech Communication Association (Interspeech)* (pp. 11-15).
12. Ponce-Lopez, V., Chen, B., Oliu, M., Corneanu, C., Claps, A., Guyon, I., & Escalera, S. (2016). ChaLearn LAP 2016: First Round Challenge on First Impressions-Dataset and Results. In *Computer Vision-ECCV 2016 Workshops* (pp. 400-418). Springer International Publishing.
13. Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1), 1929-1958.
14. Verhoeven, B., Daelemans, W., & De Smedt, T. (2013). Ensemble methods for personality recognition. In *Proceedings of the Workshop on Computational Personality Recognition* (pp. 35-38).