Predicting User Psychological Characteristics from Interactions with Empathetic Virtual Agents

Jennifer Robison[1], Jonathan Rowe[1], Scott McQuiggan[2] and James Lester[1]

[1] Department of Computer Science, North Carolina State University, Raleigh, North Carolina
[2] SAS Institute, Cary, North Carolina

Abstract. Enabling virtual agents to quickly and accurately infer users' psychological characteristics such as their personality could support a broad range of applications in education, training, and entertainment. With a focus on narrative-centered learning environments, this paper presents an inductive framework for inferring users' psychological characteristics from observations of their interactions with virtual agents. Trained on traces of users' interactions with virtual agents in the environment, psychological user models are induced from the interactions to accurately infer different aspects of a user's personality. Further, analyses of timing data suggest that these induced models are also able to converge on correct predictions after a relatively small number of interactions with virtual agents.

Keywords: User modeling and adaptive agents, Evaluation of virtual agents, Agents in narrative, Affective interaction

1 Introduction

A central feature of social intelligence is the ability to infer psychological characteristics of an interlocutor from a relatively limited number of observations. Quickly sizing up another participant in social settings is both common and useful in human-human interactions. Providing virtual agents with the ability to infer users' psychological characteristics could contribute to increased agent believability. This, in turn, could contribute to greater immersion in narrative games [1, 2] and increased effectiveness of game-based learning environments [3, 4] and agent-based training systems [5]. Alternatively, the same information could be obtained by requiring users to fill out lengthy questionnaires or to be extensively questioned by virtual agents. This approach, though frequently used, is invasive and for some applications may be undesirable or infeasible. Enabling a virtual agent to quickly and accurately infer users' psychological characteristics such as their personality could support a broad range of applications in education, training, and entertainment.

This paper presents an inductive framework for inferring users' psychological characteristics from observations of their interactions with virtual agents in a narrative-centered learning environment. Psychological user models are trained on traces of users' interactions with virtual agents in the environment. As users perform tasks, they engage in pre-scripted branching dialogue-like interactions with multiple virtual agents. Three families of attributes are monitored and logged: situational attributes (users' locations, tasks, actions), affective attributes (user-reported emotion states), and conversational attributes (frequency and duration of user-agent interactions). Psychological user models are then induced from the interactions to infer various aspects of user personality (based on Big Five models of personality [6]). Further, analyses of timing data suggest that these induced models are able to rapidly converge on correct predictions after a relatively small number of interactions with virtual agents.

2 Related Work

Recent work on intelligent virtual agents has investigated a range of models for the expression of virtual characters' social behaviors. Gratch *et al.* found that virtual agents exhibiting contingent non-verbal feedback could effectively engender feelings of rapport while listening to human speakers [7]. Other work has sought to construct computational models of socially normative conversational behavior [8, 9]. The characterization of different social relationships that can exist between humans and virtual agents has also received considerable attention. In their seminal work, Reeves and Nass asserted that humans treat computers as social actors, and the same social rules that govern human-human interactions also apply to human-machine interactions [11]. This has led to several lines of investigation examining the social-cognitive impacts of different factors on short-term and long-term human-agent interactions [12, 13].

The ability to automatically attribute traits to others, to assign stereotypes and to perform related inferences almost instantaneously, is a powerful and pervasive social phenomenon [11]. This ability to quickly "size up" another individual provides humans valuable knowledge about how to proceed during an interaction. Some recent work on social inference in virtual agents has focused on developing and evaluating a computational framework of social causality [14], but to our knowledge, little work to date has focused on automatically inferring users' psychological characteristics, such as personality.

3 Experimental Method

In this study, personality profiles of human subjects and traces of subjects' interactions with virtual agents are used to learn predictive models of user personality. These interactions took place within CRYSTAL ISLAND, a narrative-centered learning environment targeting eighth-grade (12-13 year old) students studying microbiology. Within this environment, students are attempting to solve a science mystery by gathering information from objects and interactions with virtual characters. For more information on the CRYSTAL ISLAND environment, please see [4, 15]

3.1 User Study

To investigate the inductive framework for inferring users' psychological characteristics from interactions with empathetic virtual agents, a human participant study was conducted with 35 college students ranging in age from 21 to 60 (M = 24.4, SD = 6.41) including 9 females and 26 males. Human-agent interaction histories acquired from monitoring and logging these subjects' interactions with virtual agents in the CRYSTAL ISLAND learning environment offer a rich substrate for developing empirically grounded models of cognition, affect, and social behavior. Previously they have been used to study empathy in virtual agents [15] and affect in students [3], and here they provide training data for inducing predictive models of user psychological characteristics.

Though many psychological characteristics may be of interest, this study focused on learning models of user's personality. Individuals' *personalities* are dispositions over a long duration, in contrast to emotions or moods which are more limited in their duration [16]. The Big 5 Personality Questionnaire [6] was used to assess subjects' personality. It employs five principal categories: *openness*, *conscientiousness*, *extraversion*, *agreeableness* and *neuroticism*. Subjects indicated their agreement with 45 statements across each of these five dimensions using a 5 point Likert scale [6]. Each of these measures may impact the receptiveness of a student towards various agent behaviors [17, 18].

In the study, participants were given an overview of the experiment agenda, and then completed several pre-experiment instruments, including the Big 5 Personality Questionnaire. Following completion of these materials, participants were given 45 minutes to solve the mystery. During this time, users interacted with each of six virtual agents who inhabit CRYSTAL ISLAND, each playing a distinct role in the environment. For consistency, each time a subject decided to interact with a virtual agent, a set schema was used to provide an empathetic interaction. Students were first greeted with an introduction of "Hi Alex, how are you feeling?" to which students were able to respond by selecting one of 10 emotional states (*anger, anxiety, boredom, confusion, delight, excitement, fear, flow, frustration and sadness)*. The characters then responded with a short, one or two sentence empathetic response, followed by content related to the narrative environment. At the end of the interaction, characters would conclude by asking students "How are you feeling now?" to which students could respond using the same set of emotions.

3.2 Inductive Framework

As subjects interacted with the virtual agents in the learning environment, all of their interactions were monitored and logged. In addition to "snap-shot" attributes that offer a view onto the interaction at a particular moment, the value of "higher-order" attributes that summarize interactions up to that moment were also periodically computed, e.g., the number of interactions that a subject has had with all of the agents up to that point in time. Three categories of attributes were used to induce predictive models for user psychological characteristic modeling: situational attributes, affective attributes, and conversational attributes. *Situational attributes* are extracted directly from behavior trace data, and include users' locations within the environment, their tasks, and the actions they take to perform those tasks. *Affective attributes* feature higher-order attributes derived from users' self-report responses to agents' inquiries about how they feel, as well as the relative frequency of each emotion reported up to the current moment. *Conversational attributes* record the frequency and duration with which users communicate with the environment's virtual agents.

From these three categories of attributes, four separate datasets were created. Dataset 1 (*Situational-Only*) consisted solely of situational attributes taken directly from the user logs. Dataset 2 (*Situational/Affective*) and Dataset 3 (*Situational/Conversational*) supplemented the situational attributes with either affective or conversational attributes, respectively. Dataset 4 (*All*) included all three types of attributes.

4 Results

For each trait (e.g., *extraversion*) which makes up a personality profile, each subject was classified as being either "High" or "Low" on that trait by splitting on the median value for that trait across all subjects. Since this median is likely specific to the population studied, further investigation will be required to create comprehensive models applicable to a variety of populations. Using the behavior traces of the subjects interacting with the virtual agents as training data, predictive models were learned for each individual trait and then aggregated to form models of personality. Using the WEKA machine learning toolkit [19], three types of models were learned for each of the 5 traits using each of the four datasets above: Naïve Bayes, Decision Tree, and Support Vector Machines. All models were learned using a tenfold cross-validation scheme for producing training and testing datasets. In this scheme, data is decomposed into ten equal partitions, nine of which are used for training and one of which is used for testing. The equal parts are repeatedly swapped between training and testing sets until each partition has been used for both training and testing. Tenfold cross-validation is widely used for obtaining an acceptable estimate of error [19].

The accuracy of aggregated models for each dataset and model type are presented in Table 1. All of the models outperformed baseline measures with statistical significance. The baseline for each trait is computed by choosing the most frequent class label and applying it to all test data. The baselines for each specific trait are then aggregated to provide the average baseline measure (54.4%) for overall personality (high agreeableness, conscientiousness, extraversion and openness – low neuroticism). For instance, the worst performing model that was learned, which was the naïve Bayes model of *concientiousness*, used only situation attributes but was still significantly higher than the baseline (56.02%, p<0.05). The model with the highest accuracy was the decision tree model of *agreeableness*, which was learned from situation, agent and affect-based attributes (96.9%). Overall, decision tree models outperformed support vector machines (M_1=84.9%, SD_1=15.7%, M_2=73.6%, SD_2=11.2%, p<0.0001), which in turn outperformed naïve Bayes models (M=67.0%, SD=8.7%, p<0.0001).

Table 1. Accuracies by model type and dataset (clockwise from top-left: Datasets 1, 2, 4, 3).

	Without Affective Attributes		With Affective Attributes	
Without Conversational Attributes	NB	57.8%	NB	72.5%
	DT	58.6%	DT	93.4%
	SVM	58.1%	SVM	79.4%
With Conversational Attributes	NB	60.0%	NB	73.6%
	DT	91.3%	DT	96.4%
	SVM	73.8%	SVM	84.0%

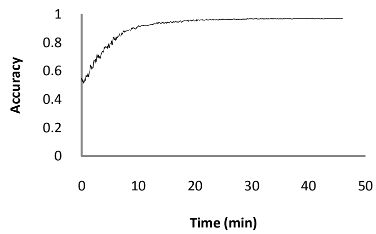

Figure 1. Speed of convergence on accurate personality traits.

Models also varied significantly based on the attributes used to train them. The inclusion of affect attributes (the *Situational/Affective* Dataset) increased the accuracy of user psychological characteristics predictions over the situational attributes alone (the *Situational-Only* Dataset) (M_1=81.7%, SD_1=10.1%, M_2=58.1%, SD_2=1.1%, p<0.0001). The inclusion of conversational attributes (the *Situational/Conversational* Dataset) also increased the accuracy of user psychological characteristics predictions over the situational attributes alone (the *Situational-Only* Dataset) (M_1=75.0%, SD_1=14.3%, M_2=58.1%, SD_2=1.1%, p<0.0001). Overall, affect information provided a larger increase in performance than conversational information (p=0.14). However, the largest performance increase came from the inclusion of these two sets of attributes together, which provided a slight increase in performance over affective data alone (M_2=84.3%, SD_2=10.7%, M_2=81.7%, SD_2=10.1%, p<0.50).

Because the overarching objective of this work is not only to infer a user's psychological characteristics after observing her for an extended period of time, but also to predict these characteristics at runtime so that agents can utilize the information to guide their social behaviors with the user, it is important to understand how quickly each individual's traits can be accurately predicted. Models that quickly and accurately converge on the correct prediction are highly desirable. To assess convergence properties, the models with the highest accuracy (decision tree models trained with situational, affective, and conversational attributes) were used to create incremental estimations of users' personality traits.

To simulate runtime estimation, we applied the model to the existing logs of the users' interactions at moments when new information was obtained. These estimates provided a best guess, given current knowledge, of each of the user's traits. Over time the history of these guesses is accumulated to provide confidence estimates for the classification of each trait. The chart presented in Figure 1 shows the convergence of estimates on the accurate classification as a function of time. These results indicate that models reach a 75% confidence in the correct classifications of user traits in approximately 4 minutes of interaction time. Given 10 minutes, predictive accuracy climbs to 90%, and with 19 minutes, 95% accuracy is achieved. With character interactions taking place, on average, approximately every 4 minutes, we can see that user personality traits can be learned in very few interactions with characters.

The results indicate that interactions with empathetic virtual agents can be used to accurately infer users' overall personality as well as each of their individual traits. For all of these traits, the inclusion of both affective and conversational attributes yielded the most accurate models. This suggests that the more information agents have about the user's situation, emotions and conversational interactions, the more likely they will be to effectively adapt their behavior to suit individual users. While both affective and conversational attributes were useful for informing models, the largest gain in insight was gleaned from the affective information, suggesting that in the case of limited computational resources, this mode of data collection would prove to be the most useful. In addition to being able to learn models of personal traits, the convergence analysis of these models reveals that they can be applied and used with an individual user and within a relatively short amount of time.

5 Conclusions and Future Work

Humans have the ability to quickly assess the personality and dispositional traits of those around them through social interaction. In virtual agents, these capabilities can be provided by empirically grounded models that are induced from observations of human-agent interactions. Induced models show promise for enabling agents to accurately infer user psychological characteristics such as personality, as well as the individual traits that make up an individual's personality profile. Initial results indicate that models trained on a pool of representative users can relatively quickly converge on accurate predictions during an evolving interaction with a particular user.

The results suggest important directions for future work. For example, it will also be interesting to explore the impact of enriching the interactions between agents and users by adding less constrained, multimodal communication functionalities. It will be important to explore how incorporating the predictive models into virtual agents can be leveraged to improve agent believability and effectiveness.

Acknowledgements

The authors would like to thank the other members of The IntelliMedia Group at North Carolina State University for useful discussions and support. We are grateful to Omer Sturlovich and Pavel Turzo for use of their 3D model libraries, and Valve Software for access to the Source™ engine and SDK. This research was supported by the National Science Foundation under Grants REC-0632450, IIS-0757535, DRL-0822200 and IIS-0812291. This material is based upon work supported under a National Science Foundation Graduate Research Fellowship. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. Mateas, M., Stern, A.: Structuring Content in the Façade Interactive Drama Architecture. In: First Conference on Artificial Intelligence and Interactive Digital Entertainment, pp. 93–98. AAAI Press, Menlo Park, CA (2005)

2. Mott, B., Lester, J.: U-Director: A Decision-Theoretic Narrative Planning Architecture for Storytelling Environments. In: 5th International Conference on Autonomous Agents and Multiagent Systems, pp. 977-984 (2006)
3. McQuiggan, S., Robison, J., Lester, J.: Affective Transitions in Narrative-Centered Learning Environments. In: 9th International Conference on Intelligent Tutoring Systems, pp. 490-499. (2008)
4. McQuiggan, S., Rowe, S., Lester, J.: The Effects of Empathetic Virtual Characters on Presence in Narrative-Centered Learning Environments. In: 26th Annual SIGCHI Conference on Human Factors in Computing Systems, pp. 1511–1520. ACM Press, New York (2008)
5. Johnson, W. L.: Serious Use of a Serious Game for Language Learning. In: 13th International Conference on Artificial Intelligence in Education, pp. 67–74. IOS Press (2007)
6. McCrae, R., Costa, P.:Personality in Adulthood: A Five-Factor Theory Perspective, Second Edition. Guilford Press, New York (2003)
7. Gratch, J., Wang, N., Gerten, J., Fast, E., Duffy, R.: Creating Rapport with Virtual Agents. In: Pelachaud, C., Martin, J.-C., Andre, E., Chollet, G., Karpouzis, K., Pele, D. (eds.) IVA 2007. LNCS (LNAI), vol. 4722, pp. 223–236. Springer, Heidelberg (2007)
8. Si, M., Marsella, S., Pynadath, D.: Thespian: Modeling Socially Normative Behavior in a Decision-Theoretic Framework. In: Gratch, J., Young, M., Aylett, R., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS (LNAI), vol. 4133, pp. 369–382. Springer, Heidelberg (2006)
9. Pedica, C., Vilhjalmsson, H.: Social Perception and Steering for Online Avatars. In: Prendinger, H., Lester, J., Ishizuka, M. (eds.) IVA 2008. LNCS (LNAI), vol. 5208, pp. 104-116. Springer, Heidelberg (2008)
10. Reeves, B., Nass, C.: The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press, New York (1996)
11. Uleman, J.S.: A Framework for Thinking Intentionally About Unintended Thoughts. In: Uleman, J.S., Bargh, J.A. (eds.), Unintended Thought, pp. 425–449. Guilford Press, New York (1989)
12. Bickmore, T. and Picard, R.: Establishing and Maintaining Long-Term Human-Computer Relationships. ACM Transactions on Computer-Human Interaction 59(1), 21–30 (2005)
13. Kim, Y., Baylor, A.: A Social-Cognitive Framework for Pedagogical Agents as Learning Companions. Educational Technology Research & Development 54(6), 569–590, (2006)
14. Mao, W., Gratch, J.: Modeling Social Inference in Virtual Agents. Journal of Artificial Intelligence and Society. Journal of Artificial Intelligence and Society 24(1), 5–11 (2009)
15. McQuiggan, S., Robison, J., Phillips, R., Lester, J.: Modeling Parallel and Reactive Empathy in Virtual Agents: An Inductive Approach. In: 7th International Joint Conference on Autonomous Agents and Multi-Agent Systems, pp. 167–174. (2008)
16. Rusting, C.: Personality, Mood, and Cognitive Processing of Emotional Information: Three Conceptual Frameworks. Psychological Bulletin124(2), 165-196 (1998)
17. Isbister, K., Nass, C.: Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. International Journal of Human Computer Interaction Studies 53,251–267 (2000)
18. Kang,S.,Gratch,J.,Wang,N.,Watts, J.: Agreeable People Like Agreeable Virtual Humans, In: Prendinger, H., Lester, J., Ishizuka, M. (eds.) IVA 2008. LNCS (LNAI), vol. 5208, pp. 253–261. Springer, Heidelberg (2008)
19. Witten, I., Frank, E.: DataMining: Practical Machine Learning Tools and Techniques, Second Edition. MorganKaufmann, San Francisco (2005)