Int. J. Human-Computer Studies 72 (2014) 567–583

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

A framework for the assessment of synthetic personalities according to user perception ☆

Zoraida Callejas [a,*], David Griol [b], Ramón López-Cózar [a]

[a] Department of Languages and Computer Systems, University of Granada, CITIC-UGR, Pdta. Daniel Saucedo Aranda s/n, 18071 Granada, Spain
[b] Department of Computer Science, Carlos III University of Madrid, 28911 Leganés, Spain

ARTICLE INFO

Article history:
Received 12 April 2013
Received in revised form
18 January 2014
Accepted 16 February 2014
Communicated by Pelachaud Catherine
Available online 15 March 2014

Keywords:
Personality
Evaluation
Conversational agents
Human–computer interaction
Human–robot interaction

ABSTRACT

Endowing artificial conversational agents with personality is a very promising way to obtain more believable user interactions with robots and computers. However, although many authors have studied how to create an agent's personality and how it affects performance and user satisfaction, less attention has been paid to assess whether the designed agent's personality corresponds to the users' perception, whether it is easily recognizable, and what is the effect that the user's own personality has in the discrimination of the agents' personality. In this paper we present an assessment framework to address these issues in an integrated way, which in our opinion offers enough flexibility to consider the diversity of application domains and evaluation approaches that can be found in the literature. The framework is based on numerical measures, which facilitate the interpretation of results and makes it possible to compare and rank different agents with respect to the user's perception of the rendered personality. In addition, we have developed a tool that implements the framework, which may be very useful for researchers in order to easily evaluate different agent personalities.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Artificial conversational characters implemented in robots or computers play an increasingly prominent role nowadays, as they offer an enhanced user experience and robust transmission of information through multimodal interaction and personalized content. These agents are used in many different domains with varying complexities, such as education (Kerly et al., 2007; Veletsianos and Miller, 2008; Wik and Hjalmarsson, 2009), commerce (Lucente, 2000; Shimazu, 2002; Semeraro et al., 2008), healthcare and training (Bickmore and Schulman, 2007; Nouza et al., 2011; Turunen et al., 2011), elderly care (Sharkey and Sharkey, 2011; Sakai et al., 2012; Sharkey and Sharkey, 2012), smart environments (Ábalos et al., 2010; Bien et al., 2008; Park et al., 2008; Bezold and Minker, 2010; Carolis et al., 2010; Griol et al., 2013) and personal company (Benyon and Mival, 2007; Kerly et al., 2007; Castellano et al., 2010; Decker et al., 2011; Kramer et al., 2011; Pesty and Duhaut, 2011; Danilava et al., 2012).

It is very important to make the agents usable, engaging, and effective in order to foster user commitment with them. To achieve this goal, the agents must have believable interaction capabilities, which would evoke social responses from the users (Ochs and Pelachaud, 2013) and improve task performance (Simmons et al., 2011). Cassell and Tartaro (2007) provide a very interesting discussion on the difference between believable agents and believable interactions. It has been argued that user satisfaction with the agents depends more on whether the interaction is realistic or not (von der Pütten et al., 2010), than on whether their appearance is realistic or not, even though their shape has an effect on users (Hwang et al., 2013).

Believable interactions imply that users can apply their models of human communication to the agents, which must fulfil the users' expectations and have a convincing and intuitive behaviour (Schonbrodt and Asendorpf, 2011a). A key to such behaviour is an appropriate integration of verbal and non-verbal cues (Salem et al., 2011), sociocultural context, emotional responses and long-term attitudes toward each user (Gockley et al., 2006; Mower et al., 2009; Simmons et al., 2011). Thus, it is important that these agents become recognizable individuals in order to have life-like interaction capabilities (Cañamero et al., 2006), which makes personality a key aspect to be addressed when designing these agents.

In Psychology, personality defines the recognizable behavioural style of a person, which is reflected in their responses to perception, learning, attending, remembering, problem-solving, and expressing emotions (Özakpinar, 2010). Several studies have shown that users

☆This paper has been recommended for acceptance by Pelachaud Catherine.

* Corresponding author. Tel.: +34 958 241 000x20049; fax: +34 958 243 179.
E-mail addresses: zoraida@ugr.es (Z. Callejas), dgriol@inf.uc3m.es (D. Griol), rlopezc@ugr.es (R. López-Cózar).

http://dx.doi.org/10.1016/j.ijhcs.2014.02.002
1071-5819/© 2014 Elsevier Ltd. All rights reserved.

are ready to assign personality to synthetic characters (Nass and Lee, 2000). Hence, there is a great interest in finding appropriate models to render consistent personalities that enable believable interactions with the agents. Mairesse and Walker (2010) propose to tailor the agent's personality according to the application domain. For example, in a tutoring system they suggest to render extrovert and agreeable pedagogic agents, whereas it could be interesting for a psychotherapy agent to be neurotic. They also point out that the personality rendered by telesale agents could match the company's brand.

Other studies have focused on how to adapt the agent's personality to match users' personality. Goren-Bar et al. (2006) discussed that adaptivity must be fine tuned to personality, as they found that neurotic people tend to reject social technology and people who have good control of their own impulses tend to dislike content adaptation on the basis of interest. This is why other authors have taken into account personality as well as interests and tastes to provide personalized information and services, as in Pu et al. (2012). Several works in evolutionary and social psychological theories of inter-personal attraction provide empirical evidence that similarity can be a determinant of attraction and plays a key role in the development of human relations such as familiarity friendship, group affiliation or romantic relationships (Klohnen and Luo, 2003; Montoya and Horton, 2013). This phenomenon is known as *the similarity attraction hypothesis* (Byrne and Nelson, 1965).

As will be discussed in Section 2, different authors have demonstrated that users are more attracted to agents with behaviours similar to their own. For example, Endrass et al. (2011) developed culture-specific small talk dialogues for virtual agents in German and Japanese and demonstrated that virtual agents whose behaviour was influenced by a cultural background were evaluated more positively and that users preferred dialogues in line with the observations made for their own cultural background. Other authors have also demonstrated that users prefer synthetic agents that are similar to them either in static features such as age, gender or ethnicity (Qiu and Benbasat, 2010), in dynamic features such as prosody (De Looze et al., 2014), at a behavioural level as for example feedback-giving style in online training (Behrend and Thompson, 2011), or when their personality is similar to their own. For example, Nass and Lee (2000) and Nass and Yen (2012) showed that users' perception of the agent's intelligence and competence increases if the perceived agent's personality matches their own. However, other authors claim for the need to study similarity or complementary attraction in the light of the application domain and the users' expectations (Joosse et al., 2013).

Personality models are usually created ad hoc and the results obtained by different authors are difficult to compare. Also, a clearly defined way of evaluating the extent to which a specific model succeeds in synthesizing a recognizable personality does not exist. There are different evaluation frameworks for multimodal systems, such as PARADISE (Walker et al., 1997; Hajdinjak and Mihelic, 2006), which were created with the aim of providing a common error measurement for analysis, interpretation and comparison. However, they are focused on the evaluation of communication, and although they can be used to measure the effect of personality on user–agent communication (e.g., user satisfaction, and communication performance and success), they cannot be used to explicitly assess personalities.

In fact, the great majority of studies that have addressed the assessment of personality have been carried it out in terms of the acceptance and/or performance of the agent based on the rendered cues (e.g., to study whether a extrovert agent improves acceptance), and studies in which the perceived personality is compared to other alternative personalities (e.g., to study whether an extrovert personality is preferred by users rather than an introvert personality). Less attention is paid to the actual perception of the rendered personality (e.g., whether the users really perceive the agent as extrovert).

In this paper we contribute an assessment framework that constitutes a novel instrument to evaluate personality in three main dimensions:

1. Whether the rendered personality is perceived by the users as the designers intended. For example, if the designers plan an extrovert personality, whether users perceive it as extrovert or as something else.
2. Whether the personality is recognizable, that is, if users perceive it consistently (i.e. if users agree in their perceptions, or different users perceive very different personalities).
3. Whether the agent's personality matches the users' personality, and how the previous dimensions are affected by the personality of users.

The proposed framework is of general purpose and can be used for a wide range of assessments regardless of the application domain. It can deal with corpora that follow different forms and distributions. Also, it supports comparison and rank ordering of different personalities, and provides numerical results that are easy to interpret and allow for more informed decisions. Finally, it accounts for the effect of the similarity attraction hypothesis, which makes it robust to changes in the characteristics of the target users.

In addition, we contribute the *Agree!* software tool, which implements the assessment framework and provides automatic reports to evaluate synthetic personalities in the previously described dimensions.

The framework can not only be used to evaluate synthetic personalities, but it can also be employed in any generic study of personality in which different profiles are compared with a reference profile, or when it is interesting to compute the agreement rates in an annotated corpus. For example, to compute the similarity of a subject's personality with an average profile of a certain personality disorder (Widiger, 2005; Lawton et al., 2011), or in psychological studies based on observer ratings. In McCrae et al. (2007) there is a detailed survey on the various current uses of observer rating research, for instance to study the difference between the public and private self, self-reports and perceptions of spouses, siblings or acquaintances, or the differences in the personality of the same person at different times (e.g., before and after suffering from Alzheimer's).

On the other hand, the framework may be employed as well in areas not directly related to personality. For example, our tool could be used for studying emotion either to compute agreement for annotated corpora in which different items are tagged with an emotion label, or when the item is described as a vector in a space such as the 2D affective space (Schubert, 1999).

The rest of the paper is organized as follows. Section 2 describes related work on personalities for synthetic agents and addresses approaches available in the literature for evaluation, which we hope will help to show the benefits of our contribution. Section 3 presents the assessment framework and its two main evaluation procedures: score-based and tag-based assessments. Section 4 describes the *Agree!* tool, an implementation of the framework that we have developed with the aim of assisting researchers in the assessment of synthetic personalities. Finally, Section 5 presents the conclusions and suggests possibilities for future work.

2. Related work

People consistently associate different verbal and non-verbal markers with certain speaker personality characteristics (Neff et al., 2010).

Download English Version:

https://daneshyari.com/en/article/400818

Download Persian Version:

https://daneshyari.com/article/400818

Daneshyari.com