User Modeling as a goal in itself: an Artificial Companion for the Elderly

Judith Masthoff and Kees van Deemter

University of Brighton, UK

Judith.Masthoff@brighton.ac.uk
Kees.van.Deemter@itri.brighton.ac.uk

Abstract. User modeling tends to be seen as a "means to an end": something that is needed to adapt systems to users, and that we want to make as painless and effortless for the users as possible. This paper describes how user modeling can be a goal in itself, a pleasurable activity that provides users with somebody to have meaningful conversations with. The example of interactive Television as an artificial companion for the elderly is explored. In addition, it is shown how techniques for knowledge editing as put forward by the Natural Language Generation community can help to make user modeling easy and pleasant.

1 Background

The importance of social support

The elderly constitute an increasing proportion of the population in the UK and Europe. Increasingly, elderly people live alone, and have a small social network [1]. A poor social network carries an increased risk of developing dementia [2] and depression [3], and has a life-shortening effect [4]. Conversely, friendship and conversation have been shown to have a clear positive impact on health and well-being [5].

Television plays a central role in the life of an elderly person: a recent study in the UK shows that older people are the heaviest consumers of television, with the 65+ group watching a staggering average of 5 hours 14 minutes a day [6]. The same study shows that almost all of those aged 75+ watch television every day (96%). Television is watched to alleviate loneliness [7], and lonely people watch more television [8]. However, this is a vicious circle, as television watching tends to *reduce* social involvement [9]. Conversely, people who indicate being happy when engaged in social activity, tend to report being bored and unhappy when watching television [10].

Interactive TV as an artificial companion

In this paper, we introduce a possible future system called Artificial Companion for the Elderly (ACE). ACE is an attempt at turning the negative effects of TV watching around, by changing the TV from a piece of broadcasting technology into a stimulating *companion* whom you can have conversations with. The advent of digital television, and in particular its potential for interactivity provides the opportunity to

pursue this idea. It opens up the possibility of personalized, adaptive experiences delivered in such a way that the viewers' awareness of the world around them is enhanced and their general alertness is boosted, which will in turn make it easier for them to participate in meaningful interactions with other people. Our target user group is the (very) elderly user.

But surely, only a human companion can help?
There is evidence that pets alleviate depression in older adults, particularly those with minimal social support [11]. There has been quite some work on artificial pets: for example, the Sony dog, the Omron cats, the Tamagotchi, the Norns, the PlayStation "Pet in TV". These pets build up a simple repertoire of behaviors, based on how you take care of them. Artificial pets are being used in therapy [12]. Of course, a pet cannot replace a human, and neither can an interactive TV. We will, however, work towards the construction of an artificial but human-like companion, somebody to have (limited) two-way conversations with.

Conversations, not done before?
The last decades have seen a considerable amount of research on dialogue systems, based on a variety of input and output media. These systems tend to focus on performance of a concrete task, such as making travel reservations [13], or obtaining information on bank account balances [14]. Useful though systems of this kind may become, their aim is not to entertain, educate, and integrate their users. ACE, by contrast, does not support the elderly through performing a task, but through conversation. It needs to keep them entertained, show interest and curiosity, and encourage them to be actively engaged. The well-known example of the *Eliza* system (which models a Rogerian psychotherapist, [15]) shows that simple mechanisms can go a surprisingly long way. ACE will go beyond *Eliza* (and its 'chatterbot' heirs) by tapping into recent research in gerontology, personalization, human-computer interfaces (HCI) and natural language processing (NLP). A particular source of inspiration is the recent wave of work on (embodied) conversational agents, which try to mirror the way in which personality and emotion are expressed linguistically (and through speech, gesture, gait, gaze, facial expression). See [16, 17, 18] for examples.

HCI / Gerontechnology challenges.
Difficulties in handling computers and other modern technology increase with age [19]. This is partly due to a change in cognitive and perceptual abilities with age [20], and partly to the phenomenon of *technology generations*: the experience with technology in adolescence and young adulthood influences how users approach current technology [20]. A key challenge therefore is to keep interactions with the TV manageable even for elderly users. Our take on this problem will be to structure the interaction with the TV by letting ACE generate questions that the user will subsequently answer (using either speech or a remote control), in line with work on so-called conversational interfaces [21]. In addition, ACE will need to be able to cope with a large variation in perceptual and cognitive abilities. It will build on work done on "Design for all" (also called "Universal Access"), e.g. [22]. Other major issues will

be the integration of ACE within the normal TV broadcast, and the encouragement to activity of normally passive viewers.

2 Envisaged Application

ACE could support conversations in a narrow domain, such as the weather, which may be modeled thoroughly. This approach would, however, tend to lead to very limited, impersonal, conversations, which would not provide the viewers the support and entertainment we have in mind. A more promising approach would be to concentrate on a wide domain that is of great interest to many elderly users: *the daily news* [6]. If we engage the elderly user with the daily news, their awareness of the world around them and their ability to interact with their peers will be enhanced. The difficulty is that it will be impossible to enter all the relevant facts in a format that ACE can understand. To get around this problem, we will -- roughly in the spirit of the Semantic Web -- allow existing newsreel to be selectively *annotated*, based on a dedicated ontology. For each event reported, the ontology might allow, for example, specification of the *type* of the event (e.g., political/sporting/etc.), its *beginning and end* time, whether it is generally viewed as *good or bad*, and so on. A new annotation tool based on such an ontology would allow a specialist to make aspects of the daily news transparent to ACE, allowing it to ask intelligent questions about it even though its understanding is limited, e.g., 'Has an earthquake of this magnitude occurred earlier during your lifetime?' Needless to say, as much as possible of the annotation will be performed automatically (e.g., [23]).

3 Modeling the elderly viewer

To be able to relate the news to the interests of the elderly user, the television will need to build up a detailed user model. This will not be done in the usual indirect way (i.e., by monitoring the user's behavior), but by involving the elderly user in the construction of a user's knowledge base (UKB) that contains their family tree and/or other aspects of their personal history. As a result, ACE will often know the answers to the questions that it asks (e.g., the user was alive during the 1995 earthquake in Kobe), enabling it to respond intelligently. Crucially, the construction of the UKB will mimic human conversation (with ACE asking questions), aiming to be both user-friendly and entertaining. At any time, there will be many holes in the UKB, and therefore multiple questions competing to be asked. ACE will carefully select which questions it asks:

- It will not ask too many questions at any time, and relate them as much as possible to the broadcast. For instance, if a news item is related to younger people, then ACE could ask whether the viewer has children. If a news item mentions France, and if the viewer is British, then ACE could ask whether the viewer has ever visited France.
- It will mainly ask questions that make a substantial contribution to extending its knowledge base. As observed in [24], for some questions the answers are very

likely. For instance, when asking a British viewer whether they have ever visited Burundi, the answer is very likely to be 'no', and therefore the answer is not likely to add much to the UKB. Similarly, for viewers in the US the 'have you ever visited France' question might be replaced by a 'have you ever been to Europe' question. In contrast to [24], however, for ACE it is not always bad to ask questions it can guess (or even knows) the answer to. For instance, in the earthquake example mentioned above, ACE knows the viewer was alive during the Kobe earthquake, and therefore the answer to the question 'Has an earthquake of this magnitude occurred earlier during your lifetime?' should be 'yes'. Asking this question makes the viewer think, which is a good thing in a conversation. Besides, the viewer might not remember the Kobe earthquake, and a 'no' response could therefore lead to an interesting discussion. In a sense, some questions could serve to add information to the UKB about whether the viewer remembers something or not. A lack in memory could indicate a lack of interest in a topic.

- Follow-up questions could be asked depending on the viewer's response. For instance, if they have visited France, ACE could ask when was the last time they were over there. This is related to maintaining narrative flow. Our conversations do not tend to jump from topic to topic: rather we speak some time about one topic before switching to another. Topics that are salient (by just having been discussed in the broadcast, or in a previous question) will have a higher likelihood of being asked about. If the answer to the question about visiting Europe was 'no', a question could be asked about 'the viewer ever having been abroad'.
- The information already in the knowledge base will allow ACE to personalize its questions. For instance, if ACE already knows that the viewer has a granddaughter of say 16 years old, then it can ask whether it is expected that she will go to university (after a news item on university tuition fees).
- It will only ask questions that are easy to answer using the television remote control (so, questions that have a limited number of possible answers, or have a number as answer). This means that it cannot just ask the viewer a question like 'What job does your son have?' If it wants to get this information, then it will need to ask gradually over time a number of questions such as 'Does your son do manual work?' Which question it asks will depend on the likelihood of the answer, which may depend on the knowledge it has already gathered. For instance, if it knows the son studied at university, then it is less likely to start with the 'manual work' question.

Letting a user construct a model of herself can be viewed as a special case of *knowledge editing*, which is defined as the task of building up, extending and modifying a knowledge base. (The knowledge base in this case is the UKB of course.) Knowledge editing has been made easier in recent years by the advent of a new paradigm, which uses Natural Language Generation [25] to clarify the content of the knowledge base to the user and, more importantly in the present setting, to offer the user options for modifying it. Probably the most elaborate version of this idea is the approach called What You See Is What You Meant (WYSIWYM), developed by Richard Power and colleagues at the University of Brighton [26]. WYSIWYM has

been applied across a range of applications, including Question Answering (in a domain of Maritime Law) [27] and Document Authoring (in a pharmaceutical domain among other things) [28]. The idea of WYSIWYM is to express the content of the knowledge base in natural language text, and to allow users to modify the knowledge base by interacting with this text. In the case of our application, for example, the UKB could contain the information that Mrs. Smith has a son (∃x: Son(MrsSmith,x)), which a language generator might simply express as 'You have a son'. (Note that the same information would be expressed differently when said to someone who is not the parent.) A sentence of this kind, which expresses information in the UKB, is called a feedback text. Since sons have birth dates, the system can now offer the user the choice of entering the birth date of the son. We intend to go beyond this "standard" use of WYSIWYM in the following ways:

- Following [29], options for modifying the UKB will be offered in the form of questions to the user. For example, instead of offering the user a menu-based interface for specifying the son's birth date, the system will ask a question, e.g., 'In what year was your son born?' The user can then use the remote control to enter the year whereupon the system can use its understanding of the situation to interpret the response (e.g., '57' will be interpreted as 1957).
- Although experiments are confirming that WYSIWYM makes knowledge editing easier, the present application requires more than ease of use: to achieve the goals of section 1, knowledge editing has to become pleasurable and stimulating. For this reason, any information expressed by the system needs to be expressed colloquially, and preferably in a style mirroring that of the user (e.g., 'When was your son borne? In what year was that?'). This can be done by exploiting the UKB, which can contain information about the user's dialect, personality, etc, in the same way as it contains information about their family tree. Thus, the UKB will influence the form as well as the content of the things that the system says.

The resulting system would have some important similarities with recent Dialogue Systems, which allow users to enter queries by allowing them to answer questions [30]. The common element here is to give the system more initiative than was done by older Dialogue Systems, thereby bypassing the need for (error prone) Natural Language Interpretation. A difference with most current Dialogue Systems, however, is that ACE would tailor its output (such as the linguistic expression of style, personality and emotion) to a specific user. Experiments will have to determine the optimal 'mix' of modalities: current WYSIWYM systems use written text throughout, but it may well prove more effective, in the present application, to use speech when asking questions of the user (cf. [30]), especially if the user is visually impaired. Similarly, speech input might prove more effective than input via remote control.

4 Conclusions

Traditionally user modeling has been used to *enable* adaptation of content, presentation and interaction (nine different *purposes* of user modeling are mentioned

in [31]). For instance, Adaptive Hypermedia and Recommender systems (as used in Electronic Program Guides) rely on user modeling. Smart approaches to user modeling have been developed, involving stereotypes, observation of user actions, and explicit user feedback. Having *conversations* to build up a user model is not new in itself: it has, for instance, been applied in [32] to improve the interaction with a system that recommends restaurants. Systems of this kind perform user modeling to make the interaction with the user (e.g., the making of the restaurant recommendation) as effective as possible. We propose to do the opposite, using the conversation (and hence the user modeling) is a goal in itself: pleasant conversations are the goal. As a consequence, the domain has to be a lot wider, and natural language generation has to be more expressive than in more conventional approaches.

The application of an Artificial Companion for the elderly has been introduced as an example of such a new class of applications. It has been indicated how the naturalness of the conversations can be achieved by using a combination of natural language generation techniques and personalization.

We have only just started this work, and the user modeling and its user interface will need to be specified in more detail. Additionally, it will be vital to empirically establish the impact of ACE. In particular, it needs to be investigated whether ACE is easy enough to have a conversation with, whether it really feels like having a conversation with a person (e.g., involving a version of the Turing test), whether the conversation is on a deep enough level, and whether the interaction has the long-term effects that we are aiming at (i.e., enhanced mental activity and social involvement).

References

1. Knipscheer, C.P.M., Jong-Gierveld, J. de, Tilburg, T.G. van, and Dijkstra, P.A. (Eds.) (1995). Living arrangements and social networks of older adults. Amsterdam: VU University press.
2. Fratiglioni L., Wang H.X., Ericsson K., Maytan M., and Winblad B. (2000). Influence of social network on occurrence of dementia: A community-based longitudinal study. The Lancet 355(9212): pp 1315-1319.
3. Roberts, R. E. (1997). Prevalence and correlates of depression in an aging cohort: The Alameda County Study. Journal of Gerontology--Series B: Psychological Sciences & Social sciences, 52B, pp252-258.
4. Blazer, D.(1982). Social support and mortality in an elderly community population. American journal of epidemiology, 115. pp 684-694.
5. Carter, P., and Everitt, A. (1998). Conceptualising practice with older people: Friendship and conversation. Ageing and Society, 18 (1), pp79-99.
6. Hanley, P. (2002). The numbers game: Older people and the media. Independent Television Commission publication.
7. Rook, K..and Peplau, L. (1982). Perspectives on helping the lonely. In L. Peplau & D. Perlman (Eds.), Loneliness: A sourcebook of current theory, research and therapy. New York: Wiley.
8. Canary, D., and Spitzberg, B. (1993). Loneliness and media gratification. Communication research, 20, pp800-821.

9. Brody, G. (1990). Effects of television viewing on family interactions: An observational study. Family relations, 29, pp216-220.
10. Kubey, R, and Csikszentmihalyi, M. (1990). Television and the quality of life: How viewing shapes everyday experience. Hillsdale, NJ: Lawrence Erlbaum.
11. Garrity, T., Stallones, L., Marx, M., Johnson, T. (1989). Pet ownership and attachment as supportive factors in the health of the elderly. Anthrozoos, 3 (1), pp35-44.
12. Drexler, M. (1999). Pet robots considered therapy for elderly. http://www.cnn.com/TECH/ptech/9903/25/robotcat.idg/
13. Potamianos, A. Ammicht, E. and Kuo, H-K. (2000) Dialogue management in the Bell Labs communicator system. In proceedings of ICSLP-2000.
14. Bagga, A., Stein, G., Strzalkovski, T. (2000). FidelityXPress: a multi-modal system for financial transactions. Procs. of 6th conf. On Content-Based Multimedia Information Access.
15. Weizenbaum, J. (1965) A computer program for the study of natural language communication between man and machine. Communications of the ACM, 9, pp36-45.
16. Fleischman, M. and Hovy, E. (2002). Towards emotional variation in speech-based natural language generation. In Proceedings of the second Int. Natural Language Generation Conference (INLG-02), July. Harriman, NY, USA
17. Walker, M.A., Cahn, J.E, and Whittaker, S.J. (1996). Linguistic style improvisation for lifelike computer characters. In Proceedings of the AAAI Symposium on AI, ALife, and Entertainment. Portland.
18. De Carolis, B., Carofiglio, V, Pelachaud, C., and I.Poggi (2001). Interactive information presentation by an embodied animated agent. In Proceedings of the International Workshop on Information Presentation and Natural Multimodal Dialogue. Verona, Italy, Dec. 2001, pp 19-23.
19. Czaja, S. (1996). Aging and the acquisition of computer skills. In W.A. Rogers, A.D. Fisk, and N. Walker (Eds.), Aging and skilled performance: Advances in theory and applications, pp201-220. Mahwah, NJ: Lawrence Erlbaum.
20. Docampo-Rama (2001). Technology generations handling complex user interfaces. Ph.D. Thesis Eindhoven University of Technology. Eindhoven, The Netherlands.
21. Jellyvision. The Jack Principles. http://www.jellyvision.com/
22. Stephanidis, C., and Emiliani, P.L. (1999). 'Connecting' to the information society: a European perspective. Technology and Disability Journal, 10, pp21-44.
23. Humphreys, K. (1997). Event coreference for information extraction. Proceedings of the ACL/EACL 1997 Workshop On operational factors in practical, robust anaphora resolution. pp75-81. Madrid.
24. Rashid, A.M., Albert, I., Cosley, D., Lam, S.K., McNee, S.M., Konstan, J.A., and Riedl, J. (2002). Getting to know you: Learning new user preferences in recommender systems. Proceedings of the Intelligent User Interface conference. pp. 127-134.
25. Reiter, E. and Dale, R. (2000). Building Natural language Generation Systems. Cambridge University Press, Cambridge, UK.
26. Power, R. and Scott,D. (1998). Multilingual Authoring using Feedback Texts. In Procs. of ACL/Coling conference, Montreal.

27. Piwek, P., Evans, R., Cahill, L. and Tipper, N. (2000). Natural Language Generation in the MILE System. In Proceedings of the 'IMPACTS in NLG' Workshop, Schloss Dagstuhl, Germany, July 2000.
28. van Deemter, K. and Power, R. (In press). High-level authoring of illustrated documents. To appear in Natural Language Engineering, 9 (2), June 2003.
29. Good. A. (2001). Improving the usability of WYSIWYM. Final year student project. University of Brighton.
30. Veldhuijzen van Zanten, G., Bouma, G., Sima'an, K., van Noord, G., and Bonnema, R. (1999). Evaluation of the NLP components of the OVIS2 spoken dialogue system. In: van Eynde, Schuurman and Schelkens (Eds.), Computational Linguistics in the Netherlands 1998, Rodopi Amsterdam, pp 213-229.
31. UM97 Reader's guide. http://w5.cs.uni-sb.de/~UM97/guide.html [Accessed 5-2-2003].
32. Thompson, C.A., Göker, M.H., and Langley, P. A personalized system for conversational recommendations. http://www.cs.utah.edu/nlp/thompson-apa.pdf. [Accessed on 5-5-2003].