Dynamic User Modeling in Health Promotion Dialogs

Valeria Carofiglio, Fiorella de Rosis, and Nicole Novielli

Intelligent Interfaces, Department of Informatics, University of Bari,
Via Orabona 4, 70126 Bari, Italy
{carofiglio, derosis, novielli}@di.uniba.it
http://www.di.uniba.it/intint/

Abstract. We describe our experience with the design, implementation and revision of a dynamic user model for adapting health promotion dialogs with ECAs to the 'stage of change' of the users and to their 'social' attitude toward the agent. The user model was built by learning a bayesian network from a corpus of data collected with a Wizard of Oz study. We discuss how uncertainty in the recognition of the user's mental state may be reduced by integrating a simple linguistic parser with knowledge about the interaction context represented in the model.

1 Introduction

Advice giving in the domain of health promotion is generally based on the Prochaska and Di Clemente's Transtheoretical Model of Change [11]. According to this model, plans to apply in persuading subjects to change of attitude in a problem behavior are related to their stage of change in the process of modifying beliefs and intentions about this behavior. On the other hand, persuasion theories state that knowledge of the affective state of the addressee may contribute considerably to select an appropriate persuasion strategy [9,12,13]. Using information about stage of change and affective state of the client enables therapists to adapt their advice giving plans to the positive or negative attitude of the clients towards their problem behavior and towards the therapist herself. We aim at implementing an Embodied Conversational Agent (ECA) which simulates this situation in the domain of healthy eating. In our project, the agent 'observes' the user to infer his stage of change and attitude towards the system, to adapt dynamically its persuasion plan and style to the presumed situation. We leave out of this paper the description of the dialog simulation method we adopt, which maybe found in [2], to focus on description of the method we apply in recognizing and revising dynamically the user's mental state. Integration of several media (speech, physiological parameters, facial expressions and so on) is suggested as an effective method to recognize the affective state of users in human-computer interaction. If, however, text input is the only interaction form enabled to the user, as in our conversational system, this methods is not applicable. In this case, a possibility is to integrate text analysis of individual user 'moves' [7,8] with an interpretation process that considers the context in which the move was entered. In this paper, we describe how we adopted this method for recognizing the stage of change of the user and his attitude towards the ECA. In Section 2, we briefly introduce the theories behind our model.

J. Tao, T. Tan, and R.W. Picard (Eds.): ACII 2005, LNCS 3784, pp. 723 – 730, 2005.
© Springer-Verlag Berlin Heidelberg 2005

In Section 3 we describe how we identified the two components of the user's mental state in a corpus of advice-giving dialogs collected with a Wizard of Oz study. Section 4 outlines how we learned, from this corpus, structure and parameters of the dynamic belief network which represents the user model. We conclude the paper with a discussion in Section 5.

2 The Theories Behind Our Model

The Transtheoretical model of change [11] identifies five main steps in the progression from a *wrong* behavior to a *correct* one:

- in the *pre-contemplation* stage, subjects believe that their behavior is correct and do not want to change it, now or later;
- in the *contemplation* stage, they doubt that their behavior is acceptable, think about why they follow the bad habit, consider seriously the opportunity of changing it but do not want to commit to change it soon;
- in the *preparation* stage, they believe that their behavior should be changed and intend to do it soon;
- in the *action* stage, they are following a plan to change behavior (from some months);
- in the *maintenance* stage, they are maintaining change from more than 6 months.

Changing from one step to the next one may be more or less fast, depending on how serious is the problem, how persistent is the subject in following the wrong behavior and how persuasive are the arguments employed by the therapist to convince him. As demonstrated by the definitions above, the model suggests that the various stages may be recognized from a *set of signs*: belief that behavior is 'correct' or 'wrong'; intention to change it if wrong; belief that (internal and external) conditions exist to change it; .. and so on. It suggests, as well, actions the therapist may adopt to promote a correct behavior in every stage.

As far as the social attitude of the user towards the system is concerned, after Nass and colleagues demonstrated several forms of 'antropomorphic behavior' of users towards technologies [10], ECAs were proposed as a new paradigm which should enhance sociality of interaction. Among the affective aspects of communication, *interpersonal warmth* appears to be the most relevant in this case: this has been defined as "*the pleasant, contented, intimate feeling that occurs during positive interactions with friends, family, colleagues and romantic partners"* and is displayed by means of verbal expressions such as self-disclosure, emotional expressiveness, verbal immediacy, personal idioms and others (see an extended analysis of research in this domain, in [1]). Although we know that such a kind of intimacy cannot be established in short-term interactions of the kind we consider in our system, we expected to find at least some traces of these expressions as signs of the social attitude of users towards our ECA.

3 The Corpus of Dialogs

The ECA we employ in our advice giving dialogs is a quite realistic agent which takes the appearance of a young woman (see figure 1). It was implemented with a commercial

graphical software (Haptek) and text-to-speech synthesizer (Loquendo), after manipulating them to create new affect expressions and enable the agent to pronounce any utterance labelled with an appropriate mark-up language.

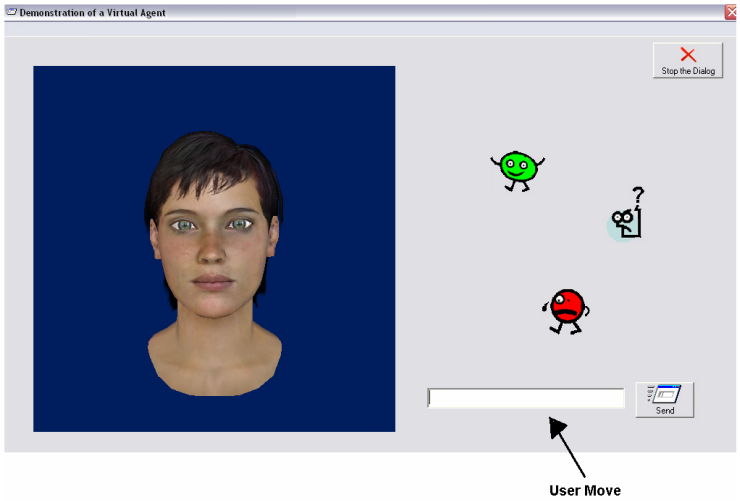

Fig. 1. The Embodied Conversational Agent employed in our Wizard of Oz studies

To test the kind of relationship users establish with our agent and the behavior the agent should show to respond appropriately to the user requirements, we performed a Wizard of Oz (WOZ) study. This study involved 30 subjects between 23 and 30 years of age, equidistributed in gender and background (graduated students in humanities or computer science). In the experiment, subjects interacted with the ECA to discuss their eating habits and received information and suggestions about possible problems discovered during the conversation. A detailed description of the WOZ tool and the experiment conditions may be found in [4]: we here synthesize its major aspects. Subjects involved in the study were initially informed about the aim of the conversational system and compiled a questionnaire assessing their attitude and behavior towards healthy eating. When ready, they asked to start the conversation with the ECA and, like in every WOZ study, thought they were interacting with a working dialog system. They could just answer the agent questions or 'take the initiative' in the dialog, by making comments and raising questions to the agent. When the subject decided to stop the dialog, a final questionnaire was displayed on the video, to collect subjective evaluation (in a Likert scale) of information provided (how much credible, persuasive, etc it was) and of the agent (how much competent, sincere etc it was). Every dialog was stored in a log at the end of interaction. The corpus of 30 dialogs and 712 subject moves was analyzed from both a quantitative and a qualitative viewpoint. We defined two measures of the subject's attitude during the dialog: a) *level of involvement*, as a function of the dialog duration and of the average length (in characters) of the user moves, and b) *degree of initiative*, as a function of the percentage of questions raised by the subject over all the dialog moves. Subjects showed several

signs of social emotions (sympathy, appreciation or irritation, disappointment) and of social relationship with the agent of the kind suggested in [1]. They were mostly in the pre-contemplation stage and a few of them changed of stage during the dialog: this is not surprising, considering that young people are interested in this topic and that unhealthy dieting is not a very serious problem. We defined a set of *signs of social attitude* of the subject towards the agent that we could recognize with linguistic analysis of the subject moves and defined a markup language with which three independent raters were asked to annotate the dialogs. Table 1 describes the markup language and shows the main results of annotation in terms of two indices: (i) *majority agreement*, as the percentage of cases in which at least two of the three raters agreed on labelling a move, and (ii) *kappa* [5], a chance-corrected measure of agreement. As the value of kappa depends on how skew is the distribution of the considered variable, given an agreement rate its value depends on the frequency of the sign of interest: this explains why the two measures are not closely correlated.

Table 1. Markup language and agreement among raters for signs of social attitude towards the agent and stage of change

Sign with definition [value]	Majority agreement	Kappa
Friendly self-introduction [Yes/No]	.98	.87
Whether the subject introduces himself with a friendly attitude (e.g. by giving his name or by explaining the reasons why he is participating in the dialog		
Familiar style [Yes/No]	.33	.16
Whether the subject employes a current language, dialectal forms, etc		
Talks about self [Yes/No]	.73	.64
Whether the subject provides information about self which was not explicitly requested by the agent		
Personal questions about the agent [Yes/No]	.70	.56
Whether the subject asks private information to the agent or tries to reciprocate suggestions		
Humor [Yes/No]	.84	.36
Whether the subject makes any kind of verbal joke in his move		
Comments on the dialog [Positive/Negative]	.82/.86	.42
Whether the subject makes any kind of comment on the agent behavior in the dialog		
Friendly farewell [Yes/No]	.93	.65
Whether the subject requests to carry-on the dialog or employs any friendly form of leave		
Believes behavior wrong [Yes/No]	.91	.46
Intends to change [Yes/No]	.97	.54

The table shows that there were good majority agreements and kappa for friendly self-introduction, talks about self and friendly farewell and a reasonably good value for personal questions about the agent. Familiar style had a quite low rate and kappa, while humor had a high rate but a low kappa. Agreement was quite good for the two signs of stage of change.

4 User Models as Dynamic Belief Networks

We applied results of the corpus analysis to build a model of the user's mental state which includes the dimensions of interest for dialog adaptation. The user modeling procedure integrates two components: (i) a very simple *parsing* of the user moves based on word semantics and salience and (ii) a *dynamic belief network* (DBN) which considers the context in which the the move was uttered, by handling uncertainty in the relationships among the various components. The model is employed to infer how the mental state of the user evolves during the dialog in relation to the dialog history. The dimensions of the user's mental state which are relevant in advice-giving dialogs are, as we said, stage of change and social attitude towards the ECA. These are *hidden* variables which depend on *observable* ones, such as the 'stable' characteristics of the users (their background and gender), the context in which the move was entered (previous agent move) and the linguistic features of the user move recognized by parsing. *Intermediate* variables are the signs of mental state listed in Table 1. We simulate the following situation: at the beginning of interaction, the model is initialized by assigning a value to the stable user characteristics. At every dialog step, knowledge about the context and evidence produced by the parser are entered and propagated in the network: the model revises the probabilities of hidden variables. As the situation gradually evolves from a dialog step to the next one, we represent the model with a DBN [6]. DBNs, also called time-stamped models, are local belief networks (called time slices) expanded over time; time slices are connected through temporal links to constitute a full model. The following is a legenda for the names of variables in figure 2:

- *Stable and known user characteristics*: background in humanities or computer science (*Back*);
- *context*: category of the previous agent move (*Ctext*), and of the current user move (*Mtype*);
- *dynamically evolving, unknown user characteristics*: social attitude towards the ECA (*Psatt*) which may be 'friendly' or 'neutral' and stage of change (*SoC*) which may be 'pre-contemplation', 'contemplation' or 'preparation'; two instances of these variables are included in the network, at times T and T+1, connected by 'temporal links';
- *signs of social attitude and stage of change*: believes behaviour wrong (*Bbw*) and intends to change (*Itc*) for *SoC*; friendly self-introduction (*Fsint*), familiar style (*Fstyl*), talks about self (*Perin*), personal questions to the agent (*Qagt*), comments on the dialog (*Comm*) and friendly farewell (*F-fw*) for the social attitude;
- *results of parsing*: these variables are in direct connection with those in the previous category: their labels start with a P.

Links among variables describe the causal relationships among stable characteristics of the users and their behavior, via intermediate nodes. DBNs, as employed in this paper, are said to be 'strictly repetitive models'. This means that structure and parameters of individual time slices is identical and temporal links between two consecutive time slices are always the same. We use a special kind of strictly repetitive model in which the Markov property holds: *the past has no impact on the future given*

the present. In our simulations, every time slice corresponds to a user move, the stable user characteristics not change from time to time (this is why we omitted the node *Back* from the figure) and temporal links are established only between dynamic subject characteristics in two consecutive time slices.

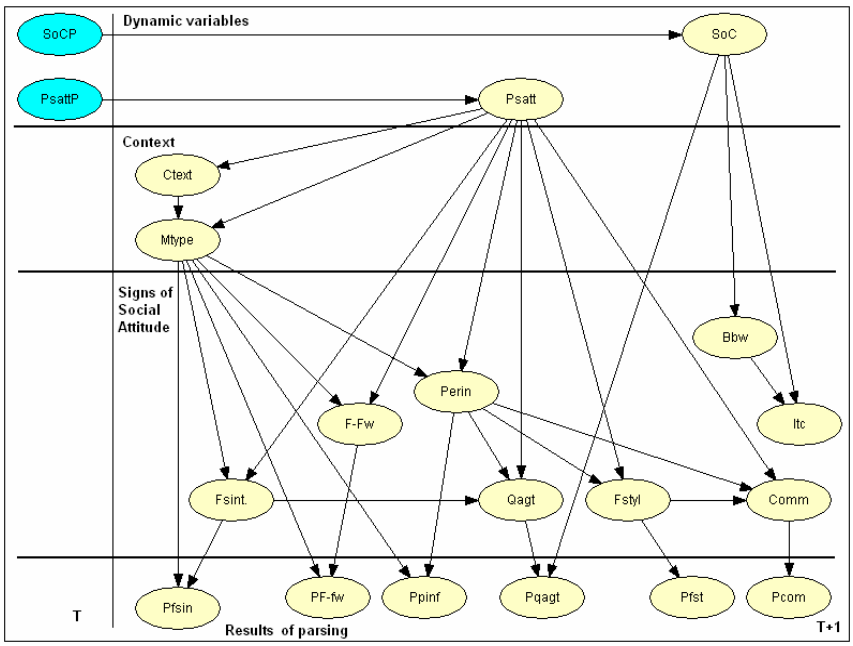

Fig. 2. The dynamic belief network representing the user model

We built the model with a combination of theoretical considerations, learning from the database of cases and subjective estimates. We employed our corpus of WOZ dialogs to train the model with the K2 learning algorithm [3]. This algorithm enables listing the nodes according to a 'search order': in our view, it is particularly appropriate in learning user models, in which links among nodes have a specified meaning and orientation, 'trigger variables' usually correspond to network roots and 'observable' ones to leaf nodes. K2 tries to find the model that fits data best by maximizing the log likelihood but does not care about the use of the resulting model and its predictive ability. We therefore applied the algorithm interactively, by learning automatically both the network structure and its parameters. At every step, we checked the plausibility of learned links, introduced, when needed, other links to avoid problems in evidence propagation due to d-separation properties and checked problems in probability tables due to 'sparse data'. The model shows that a background in humanities implies a higher level of social attitude towards the agent and friendly self introduction in particular. When the level of social attitude is low, self presentation and farewell are sometimes omitted to immediately raise questions to the agent and few comments are made after agent's answers or suggestions. The use of familiar style is associated with talking about self; this is especially included in comments or answers to the agent's

questions; personal questions about the agent prevail in the contemplation and the preparation stages of change and often come after friendly self-introduction, ... and so on. The model is employed as follows in the dialog simulation:

- at the beginning of interaction, it is initialized by propagating evidence about the user characteristics;
- after every user move, its linguistic features are analyzed with the parser and these results are introduced and propagated in the network together with evidence about the context;
- the new probabilities of the signs of social attitude and stage of change are read and contribute to formulating the next agent move, while the probabilities of 'dynamic variables' support revising high-level planning of the agent behavior.

The following is an example of simulation of the initial part of a dialog with Giulia, a subject with a background in humanities, that we did in the scope of an 'internal' validation of the predictive value of the model. A stands for 'agent' and U for 'user'; we omit from the example the analysis of SoC. The initial probability that Giulia will establish a social relationship with the ECA is $P(Psatt = Yes) = .61$.

==========

A: Hi, My name is Valentina: I'm here to suggest you how to improve your diet.

U: Hi Valentina: my name is Giulia, and I'm happy to interact with you.

Ctext = SelfIntroduction; Parsing(FriendlySelfIntroduction) = Yes; P(Pfsin) = 1; P(Fsint) = .98; P(Psatt = Yes) = .88
** After a friendly self-introduction, the probability of social attitude increases.*

A: OK: let's start, then. What did you eat at breakfast yesterday?

U: Cappuccino and cornetto

Ctext = Question; P(Psatt = Yes) = .71
** After a 'neutral' move, the probability of social attitude decreases, due to the decay effect of the link between PsattP and Psatt.*

A: What did you eat at lunch?

U: I went to a wedding where I eated a lot of fantastic food!!

Ctext = Question; Parsing(Familiar style) = Yes; Parsing(TalksAboutSelf) = Yes
P(Pfst) = 1; P(Ppinf) = 1; P(Fstyl) = .75 ; P(Perin) = .69; P(Psatt = Yes) = .88;
** The sensitivity of parsing for familiar style and talks about self is not high: hence the medium values of P(Fstyl) and P(Perin); however, the probability of social attitude after propagating this evidence and information about the context increases again.*

...

==========

5 Discussion

Health promotion dialogs are a promising application of affective computing. With the research described in this paper we studied, in particular, how dynamic user models which include an affective component may be built by integrating text analysis of dialog moves with knowledge about context and user characteristics, in a dynamically evolving user model. So far, we performed an internal validation of the predictive value of our model on some of the dialogs in the corpus: as an immediate step further, we plan to extend it to an external validation with new dialogs.

Acknowledgements. This work was financed by HUMAINE, the European Human-Machine Interaction Network of Excellence on Emotion (EC Contract 507422). We thank Berardina De Carolis and Giuseppe Clarizio for cooperating in the implementation of the WOZ tool and Loquendo for providing us their software in the scope of a scientific cooperation.

References

1. Andersen, P.A. and Guerrero, L.K.: The brightside of relational communication: interpersonal warmth as a social emotion. In Andersen, P.A. and Guerrero, L.K.: Handbook of Communication and Emotions. Research, theory, applications and contexts. Academic Press (1998).
2. Cavalluzzi, A., Carofiglio V. and de Rosis, F.: Affective Advice-Giving Dialogs. In E.André, L.Dybkjaer, W.Minker and P.Heisterkamp (Eds): Affective Dialogue Systems. Springer LNAI 3068 (2004).
3. Cooper, G.F. and Herskovitz, E.: A Bayesian method for the induction of probabilistic networks from data. Machine Learning , 9 (1992).
4. de Rosis, F. Cavalluzzi, A. Mazzotta, I. and Novielli, N.: Can embodied conversational agents induce empathy in users? AISB Virtual Social Characters Symposium (2005).
5. Di Eugenio, B. And Glass, M.: The Kappa statistics, a second look. Computational Linguistics, 2004.
6. Jensen, F.V.: Bayesian Networks and Decision Graphs. Springer (2001).
7. Lee, C.M., Narayanan, S.S., Pieraccini, R.: Combining acoustic and language information for emotion recognition. Proc. Of ICSLP (2002).
8. Litman, D., Forbes, K., Silliman, S.: Towards emotion prediction in spoken tutoring dialogues. Proc. of HLT/NAACL} (2003).
9. Miceli, M., de Rosis, F. and Poggi, I.: Emotional and non emotional persuasion. Applied Artificial Intelligence: an International Journal. (in press).
10. 10.Nass, C., Moon, Y, Fogg, B.J. and Dryer, D.C.: Can computer personalities be human personalities? Int J Human-Computer Studies, 43 (1995).
11. 11.Prochaska, J., Di Clemente, C., Norcross, H. In search of how people change: applications to addictive behavior. Americal Psychologist, 47, (1992).
12. Sillince, J.A. and Minors, R.H.: What makes a strong argument? Emotions, highly-placed values and role-playing. Communication and Cognition (1991).
13. Walton, D. The place of emotion in argument. The Pennsylvania State University Press (1992).