This document is published in:

Novais, P. et al. (eds.) (2012) *Ambient Intelligence - Software and Applications: 3rd International Symposium on Ambient Intelligence (ISAmI 2012)*. (Advances in Intelligent and Soft Computing, 153). Springer, 93-100.
DOI: http://dx.doi.org/10.1007/978-3-642-28783-1_12

Optimizing Dialog Strategies for Conversational Agents Interacting in AmI Environments

David Griol, Javier Carbó, and José Manuel Molina

Abstract. In this paper, we describe a conversational agent which provides academic information. The dialog model of this agent has been developed by means of a statistical methodology that automatically explores the dialog space and allows learning new enhanced dialog strategies from a dialog corpus. A dialog simulation technique has been applied to acquire data required to train the dialog model and then explore the new dialog strategies. A set of measures has also been defined to evaluate the dialog strategy. The results of the evaluation show how the dialog model deviates from the initially predefined strategy, allowing the conversational agent to tackle new situations and generate new coherent answers for the situations already present in the initial corpus. The proposed technique can be used not only to develop new dialog managers but also to explore new enhanced dialog strategies focused on user adaptation required to interact in AmI environments.

Keywords: Conversational Agents, Speech Interaction, Agent & Multiagent Systems for AmI, Statistical Methodologies.

1 Introduction

Ambient Intelligence (AmI) and Smart Environments (SmE) emphasize on greater user-friendliness, more efficient services support, user-empowerment, and support for human interactions. For this reason, AmI systems usually consist of a set of interconnected computing and sensing devices which surround the user pervasively in his environment and are invisible to him, providing a service that is dynamically adapted to the interaction context, so that users can naturally interact with the system and thus perceive it as intelligent.

David Griol · Javier Carbó · José Manuel Molina
Group of Applied Artificial Intelligence (GIAA), Computer Science Department, Carlos III University of Madrid
e-mail: {david.griol,javier.carbo,josemanuel.molina}@uc3m.es

To ensure such a natural and intelligent interaction, it is necessary to provide an effective, easy, safe and transparent interaction between the user and the system. With this objective, as an attempt to enhance and ease human-to-computer interaction, in the last years there has been an increasing interest in simulating human-to-human communication, employing conversational agents [4].

A conversational agent can be defined as a software that accepts natural language as input and generates natural language as output, engaging in a conversation with the user. In a conversational agent of this kind, several modules cooperate to perform the interaction with the user: the Automatic Speech Recognizer (ASR), the Language Understanding Module (NLU), the Dialog Manager (DM), the Natural Language Generation module (NLG), and the Synthesizer (TTS). Each one of them has its own characteristics and the selection of the most convenient model varies depending on certain factors: the goal of each module, or the capability of automatically obtaining models from training samples.

The application of statistical approaches to dialog management has attracted increasing interest during the last decade [8]. Statistical models can be trained from real dialogs, modeling the variability in user behaviors. The final objective is to develop conversational agents that have a more robust behavior and are easier to adapt to different user profiles or tasks. The most extended methodology for machine-learning of dialog strategies consists of modeling human-computer interaction as an optimization problem using Partially Observable Markov Decision Processes (POMDPs) and reinforcement methods. However, they are limited to small-scale problems, since the state space would be huge and exact POMDP optimization would be intractable [7].

In addition, the success of these approaches depends on the quality of the data used to develop the dialog model. Considerable effort is necessary to acquire and label a corpus with the data necessary to train a good model. A technique that has currently attracted an increasing interest is based on the automatic generation of dialogs between the dialog manager and an additional module, called the user simulator, which represents user interactions with the conversational agent [6]. The construction of user models based on statistical methods has provided interesting and well-founded results in recent years and is currently a growing research area. Therefore, these models can be used to learn a dialog strategy by means of its interaction with the conversational agent and reduce the effort to acquire a dialog corpus.

In this paper, we present a technique for learning optimal dialog strategies in conversational agents. Our technique is based on the use of a statistical dialog manager that is learned using a dialog corpus for the specific task. A dialog simulation technique is used to automatically generate the data required to learn a new dialog model. We have applied our technique to explore dialog strategies for a conversational agent designed to provide academic information. In addition, a set of specific measures has been defined to evaluate the new strategy once new simulated data is used to re-train the dialog manager. The results of the evaluation of a dialog manager developed for this agent show how the variability of the dialog model is increased by detecting new dialog situations that are not present in an initial model and selecting better system responses for the situations that were already present.

2 Our Statistical Dialog Management Technique

In most conversational agents, the conversational agent takes its decisions based only on the information provided by the user in the previous turns and its own model. This is the case with most slot-filling dialogs. The methodology that we propose for the selection of the next system answer in this kind of task is as follows [2]. We consider that, at time *i*, the objective of the dialog manager is to find the best system answer A_i. This selection is a local process for each time *i* and takes into account the previous history of the dialog, that is to say, the sequence of states of the dialog (i.e. pairs *system-turn, user-turn*) preceding time *i*:

$$\hat{A}_i = \underset{A_i \in \mathscr{A}}{\operatorname{argmax}} P(A_i | S_1, \cdots, S_{i-1})$$

where set $\mathscr{A}$ contains all the possible system answers.

As the number of all possible sequences of states is very large, we define a data structure in order to establish a partition in the space of sequences of states (i.e., in the history of the dialog preceding time *i*). This data structure, that we call Dialog Register (*DR*), contains the information provided by the user throughout the previous history of the dialog. The selection of the best A_i is then given by:

$$\hat{A}_i = \underset{A_i \in \mathscr{A}}{\operatorname{argmax}} P(A_i | DR_{i-1}, S_{i-1})$$

The selection of the system answer is carried out through a classification process, for which a multilayer perceptron (MLP) is used. The input layer receives the codification of the pair (DR_{i-1}, S_{i-1}). The output generated by the MLP can be seen as the probability of selecting each of the different system answers defined for a specific task.

3 Our Dialog Simulation Technique

Our approach for acquiring a dialog corpus is based on the interaction of a user simulator and a conversational agent simulator [3]. Both modules use a random selection of one of the possible answers defined for the semantics of the task (user and system dialog acts). At the beginning of the simulation, the set of system answers is defined as equiprobable. When a successful dialog is simulated, the probabilities of the answers selected by the dialog manager during that dialog are incremented before beginning a new simulation.

An error simulator module has been designed to perform error generation. The error simulator modifies the frames generated by the user simulator once it selects the information to be provided. In addition, the error simulator adds a confidence score to each concept and attribute in the frames. The model employed for introducing errors and confidence scores is inspired in the one presented in [5]. Both processes are carried out separately following the noisy communication channel metaphor by means of a generative probabilistic model $P(c, a_u | \tilde{a}_u)$, where a_u is the true incoming user dialog act $\tilde{a}_u$ is the recognized hypothesis, and c is the confidence score associated with this hypothesis.

On the one hand, the probability $P(\tilde{a}_u|a_u)$ is obtained by Maximum-Likelihood using the initial labeled corpus acquired with real users. To compute it, we consider the recognized sequence of words w_u and the actual sequence uttered by the user $\tilde{w}_u$.

$$P(\tilde{a}_u|a_u) = \sum_{\tilde{w}_u} P(a_u|\tilde{w}_u) \sum_{w_u} P(\tilde{w}_u|w_u) P(w_u|a_u)$$

On the other hand, the generation of confidence scores is carried out by approximating $P(c|\tilde{a}_u, a_u)$ assuming that there are two distributions for c. These two distributions are defined manually generating confidence scores for correct and incorrect hypotheses.

$$P(c|a_w, \tilde{a}_u) = \begin{cases} P_{corr}(c) & if \quad \tilde{a}_u = a_u \\ P_{incorr}(c) & if \quad \tilde{a}_u \neq a_u \end{cases}$$

4 Design of an Academic Conversational Agent

The design of our conversational agent is based on the requirements defined for a dialog system developed to provide spoken access to academic information about the Department of Languages and Computer Systems in the University of Granada [1]. To successfully manage the interaction with the users, the conversational agent carries out six main tasks described in the Introduction section: automatic speech recognition (ASR), natural language understanding (NLU), dialog management (DM), database access and storage (DB), natural language generation (NLG), and text-to-speech synthesis (TTS). The information that the conversational agent provides has been classified in four main groups: subjects, professors, doctoral studies and registration.

The semantic representation that we have chosen for the task is based on the concept of frame, in which one or more concepts represent the intention of the utterance, and a sequence of attribute-value pairs contains the information about the values given by the user. In the case of user turns, we defined four concepts related to the different queries that the user can perform to the system (*Subject*, *Lecturers*, *Doctoral studies*, and *Registration*), three task-independent concepts (*Affirmation*, *Negation*, and *Not-Understood*), and eight attributes (*Subject-Name*, *Degree*, *Group-Name*, *Subject-Type*, *Lecturer-Name*, *Deadline*,*Program-Name*, *Semester*, and *Deadline*).

The labeling of the system turns is similar to the labeling defined for the user turns. A total of 30 task-dependent concepts was defined:

- Task-independent concepts (*Affirmation*, *Negation*, *Not-Understood*, *New-Query*, *Opening*, and *Closing*).
- Concepts used to inform the user about the result of a specific query (*Subject*, *Lecturers*, *Doctoral-Studies*, and *Registration*).

- Concepts defined to require the user the attributes that are necessary for a specific query (*Subject-Name*, *Degree*, *Group-Name*, *Subject-Type*, *Lecturer-Name*, *Program-Name*, *Semester*, and *Deadline*).
- Concepts used for the confirmation of concepts (*Confirmation-Subject*, *Confirmation-Lecturers*, *Confirmation-DoctoralStudies*, *Confirmation-Registration*) and attributes (*Confirmation-SubjectName*, *Confirmation-Degree*, *Confirmation-SubjectType*, *Confirmation-LecturerName*, *Confirmation-Semester*, *Confirmation--ProgramName*, *Confirmation-Deadline*, and *Confirmation-GroupName*).

The *DR* defined for the task is a sequence of 12 fields, corresponding to the four concepts (*Subject*, *Lecturers*, *Doctoral-Studies*, and *Registration*) and eight attributes (*Subject-Name*, *Degree*, *Group-Name*, *Subject-Type*, *Lecturer-Name*, *Program-Name*, *Semester*, and *Deadline*) defined for the task.

A set of 100 dialogs was acquired with the conversational agent by means of its interaction with real users using an initial version of the system [1]. The acquisition process resulted in a spontaneous Spanish speech dialog corpus with 60 different speakers. The total number of user turns was 422. The recorded material has a duration of 150 minutes. Figure 1 shows an example of a dialog extracted from this corpus. Turns with S refer to system turns, and turns with U refer to user turns.

S1: Welcome to the university information system. You can consult information about subjects, lecturers, doctoral studies and registrations.
U1: I want to know information about subjects in Computer Science.
S2: Which subject do you want to consult?
U2: Language Processors I.
S3: Here you can consult the information of the subject. Anything else?
U3: Can you inform me about the schedules of Group D?
S4: I give the information for the Group D of Language Processors I: Monday 9:00-11:00, Wednesday 11:00-13:00. Anything else?
U4: Can you tell me the name of the lecturer?
S5: I give the information for the Group D of Language Processors I. The name of the lecturer is John Smith. Anything else?
U5: No, thank you.
S6: Thank you for using the system.

Fig. 1 Example of a dialog acquired with the interaction of the initial conversational agent with real users

5 Evaluation of Our Proposal

A dialog corpus of 3000 successful dialogs was acquired using the proposed user simulation technique following the same objectives defined for the initial acquisition with real users. A maximum number of 14 user turns per dialog was defined for the acquisition.

We have considered different dialog style features to evaluate the initial conversational agent for the task and its evolution once the simulated dialogs are incorporated to learn a new dialog model for the conversational agent. We defined and counted a set of system/user dialog acts. On the system side, we have measured the confirmation of concepts and attributes, questions to require information, and system answers generated after a database query. On the user side, we have measured the percentage of turns in which the user carries out a request to the system, provides information, confirms a concept or attribute, the Yes/No answers, and other answers not included in the previous categories. Finally, we have measured the proportion of goal-directed actions (request and provide information) versus the grounding actions (confirmations) and rest of actions.

Table 1 show the frequency of the most dominant user and system dialog acts in the initial and final conversational agents. From its comparison, it can be observed that there are significant differences in the dialog acts distribution. With regard to user actions, it can be observed that users need to employ less confirmation turns in the final agent, which explains the higher proportion for the rest of user actions using the final conversational agent. It also explains the lower proportion of yes/no actions in the final agent, which are mainly used to confirm that the system's query has been correctly provided. With regard to the system actions, it can be observed a reduction in the number of system confirmations for data items. This explains a higher proportion of turns to inform and provide data items for the final agent. Both results show that the final conversational agent carries out a better selection of the system responses.

Table 1 Percentages of different types of user dialog acts (top) and system dialog acts (bottom)

	Initial Conversational Agent	Final Conversational Agent
Request to the system	31.74%	35.43%
Provide information	20.72%	24.98%
Confirmation	10.81%	7.34%
Yes/No answers	31.47%	28.77%
Other answers	3.26%	3.48%

	Initial Conversational Agent	Final Conversational Agent
Confirmation	13.51%	10.23%
Questions to require information	18.44%	19.57%
Answers after a database query	68.05%	70.20%

In addition, we grouped all user and system actions into three categories: "goal directed" (actions to provide or request information), "grounding" (confirmations and negations), and "rest". Table 2 shows a comparison between these categories. As can be observed, the dialogs provided by the final conversational agent have a better quality, as the proportion of goal-directed actions is higher.

Finally, a total of 100 dialogs was recorded from interactions of 10 students and professors of our University employing the conversational agent developed using our proposal. We considered the following measures for the evaluation:

- Dialog success rate (*%success*). Percentage of successfully completed tasks;
- Average number of user turns per dialog (*nT*);
- Confirmation rate (*%confirm*). Ratio between the number of explicit confirmations turns (nCT) and the number of turns in the dialog (nCT/nT);
- Average number of corrected errors per dialog (*nCE*). Average of errors detected and corrected by the dialog manager;
- Average number of uncorrected errors per dialog (*nNCE*). Average of errors not corrected by the dialog manager;
- Error correction rate (*%ECR*). Percentage of corrected errors, computed as nCE/(nCE + nNCE).

Table 2 Proportions of dialog spent on-goal directed actions, ground actions and the rest of possible actions

	Initial Conversational Agent	Final Conversational Agent
Goal directed actions	62.55%	70.17%
Grounding actions	36.32%	29.59%
Rest of actions	1.13%	1.24%

Table 3 compares this acquisition with the acquisition of 100 dialogs with real users using the initial conversational agent. The results show that both conversational agents could interact correctly with the users in most cases. However, the final conversational agent system obtained a higher success rate, improving the initial results by 6% absolute. Using the final conversational agent, the average number of required turns is also reduced from 4.99 to 3.75. The confirmation and error correction rates were also improved by the final conversational agent, as the enhanced dialog model reduces the probability of introducing ASR errors. The main problem detected was related to the introduction of data in the *DR* with a high confidence value due to errors generated by the ASR not detected by the dialog manager.

Table 3 Results of the evaluation of the conversational agents with real users

	%success	nT	%confirm	nCE	nNCE	%ECR
Initial Conversational Agent	89%	4.99	34%	0.84	0.18	82%
Final Conversational Agent	95%	3.75	37%	0.89	0.07	92%

6 Conclusions

In this paper, we have described a technique for exploring dialog strategies in conversational agents. Our technique is based on two main elements: a statistical dialog methodology for dialog management and an automatic dialog simulation technique to generate the data that is required to re-train the dialog model. The results of applying our technique to the design of conversational agent that provides academic information show that the proposed methodology can be used not only to develop new dialog managers but also to explore new enhanced strategies required for the interaction in AmI environments. Carrying out these tasks with a non-statistical approach would require a very high cost that sometimes is not affordable. As a future work, we are adapting the proposed dialog management to evaluate the capability of our methodology to adapt efficiently to AmI environments that vary dynamically, in which additional information sources (including context information) must be considered in the definition of the *DR* and additional modalities and more complex operations are required for the interaction with users.

Acknowledgements. Research funded by projects CICYT TIN2011-28620-C02-01, CICYT TEC2011-28626-C02-02, CAM CONTEXTS (S2009/TIC-1485), and DPS2008-07029-C02-02.

References

1. Callejas, Z., López-Cózar, R.: Relations between de-facto criteria in the evaluation of a spoken dialogue system. Speech Communication 50(8–9), 646–665 (2008)
2. Griol, D., Hurtado, L., Segarra, E., Sanchis, E.: A Statistical Approach to Spoken Dialog Systems Design and Evaluation. Speech Communication 50(8-9), 666–682 (2008)
3. Griol, D., Sánchez-Pi, N., Carbó, J., Molina, J.: An Agent-Based Dialog Simulation Technique to Develop and Evaluate Conversational Agents. In: Proc. of PAAMS 2011. AISC, vol. 88, pp. 255–264 (2011)
4. López-Cózar, R., Araki, M.: Spoken, Multilingual and Multimodal Dialogue Systems. John Wiley & Sons Publishers (2005)
5. Schatzmann, J., Thomson, B., Weilhammer, K., Ye, H., Young, S.: Agenda-Based User Simulation for Bootstrapping a POMDP Dialogue System. In: Proc. of HLT/NAACL 2007, pp. 149–152 (2007)
6. Schatzmann, J., Weilhammer, K., Stuttle, M., Young, S.: A Survey of Statistical User Simulation Techniques for Reinforcement-Learning of Dialogue Management Strategies. Knowledge Engineering Review 21(2), 97–126 (2006)
7. Williams, J., Young, S.: Scaling POMDPs for Spoken Dialog Management. IEEE Audio, Speech and Language Processing 15(8), 2116–2129 (2007)
8. Young, S.: The Statistical Approach to the Design of Spoken Dialogue Systems. Tech. rep., Cambridge University Engineering Department (UK) (2002)