

This document is published in:

Corchado, Juan M. et al. (eds.) (2011) *Trends in Practical Applications of Agents and Multiagent Systems: 9th International Conference on Practical Applications of Agents and Multiagent Systems*, (Advances in Intelligent and Soft Computing, 90), Springer, 153-160.

DOI: http://dx.doi.org/10.1007/978-3-642-19931-8_19

© 2011 Springer-Verlag Berlin Heidelberg

Bringing Statistical Methodologies for Enterprise Integration of Conversational Agents

David Griol and José Manuel Molina

Group of Applied Artificial Intelligence (GIAA), Computer Science Department, Carlos III University of Madrid.

e-mail: {david.griol,josemanuel.molina}@uc3m.es

Abstract. In this paper we present a methodology to develop commercial conversational agents that avoids the effort of manually defining the dialog strategy for the dialog management module. Our corpus-based methodology is based on selecting the next system answer by means of a classification process in which the complete dialog history is considered. This way, system developers can employ standards like VoiceXML to simply define system prompts and the associated grammars to recognize the users responses to the prompt, and the statistical dialog model automatically selects the next system prompt. We have applied this methodology for the development of an academic conversational agent.

Keywords: Conversational Agents, Dialog Management, Statistical Methodologies, User Modeling.

1 Introduction

Conversational agents [11, 10] provide a appropriate interface between humans and computers, which is present in a growing number of commercial systems. In order to enable the rapid deployment of voice applications, markup languages such as VoiceXML have been widely adopted as they reduce the programming difficulty. However, the development of dialog systems with this approach involves a very costly engineering cycle [15]. As an attempt to reduce this cost and carry out rapid system prototyping, statistical approaches are gaining increasing interest.

There is a high variety of applications in which conversational agents can be used. One of the most wide-spread is information retrieval. Some sample applications are tourist and travel information [6], weather forecast over the phone [21], speech controlled telephone banking systems [12], conference help [1], etc. They have

also been used for education and training, particularly in improving phonetic and linguistic skills: assistance and guidance to F18 aircraft personnel during maintenance tasks [2], dialog applications for computer-aided speech therapy with different language pathologies [17].

Spoken interaction can be the only way to access information in some cases, like for example when the screen is too small to display information (e.g. hand-held devices) or when the eyes of the user are busy in other tasks (e.g. driving) [18]. It is also useful for remote control of devices and robots, specially in smart environments [13]. Finally, one of the most demanding applications for fully natural and understandable dialogs, are embodied conversational agents and companions [3, 5].

Learning statistical approaches to model the different modules that compose a dialog system has been of growing interest during the last decade [20]. Models of this kind have been widely used for speech recognition and also for language understanding [14, 9]. Even though in the literature there are models for dialog managers that are manually designed, over the last few years, approaches using statistical models to represent the behavior of the dialog manager have also been developed [16, 19]. These approaches are usually based on modeling the different processes probabilistically and learning the parameters of the different statistical models from a dialog corpus.

In this paper, we propose a statistical methodology to develop dialog managers for commercial conversational agents. Our methodology is mainly based on the modelization of the sequences of the system and user dialog acts and the introduction of a partition in the space of all the possible sequences of dialog acts. This partition, which is defined taking into account the data supplied by the user throughout the dialog, makes the estimation of a statistical model from the training data manageable. The confidence measures provided by the recognition and the understanding modules are also taken into account in the definition of this partition. The new system utterance is selected by means of a classification procedure. Specifically, we use neural networks for the implementation of this classification process. This methodology is combined with the standard VoiceXML to develop the rest of the modules of the conversational agent. We have applied our proposal to develop a conversational agent which provides academic information.

2 Our Proposal to Introduce Statistical Methodologies for the Development of Conversational Agents in Industry Applications

As stated in the introduction, our approach to integrate statistical methodologies in industry applications is based on the automatic learning of the dialog strategy using a statistical dialog management methodology. In most dialog systems, the dialog manager 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 dialog systems. The methodology that we propose for the selection of the next system

answer in this kind of tasks is detailed in [8]. Our technique is based on the definition of a data structure that we call Dialog Register (*DR*), and contains the information provided by the user throughout the previous history of the dialog. For each time i , the selection of the next system prompt A_i is carried out by means of the following maximization:

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

where set \mathcal{A} contains all the possible system answers and S_{i-1} is the state of the dialog sequence (*system-turn, user-turn*) at time i .

Each user turn supplies the system with information about the task; that is, he/she asks for a specific concept and/or provides specific values for certain attributes. However, a user turn could also provide other kinds of information, such as task-independent information. This is the case of turns corresponding to *Affirmation*, *Negation* and *Not-Understood* dialog acts. This kind of information implies some decisions which are different from simply updating the *DR*. For that reason, for the selection of the best system answer A_i , we take into account the *DR* that results from turn 1 to turn $i - 1$, and we explicitly consider the last state S_{i-1} . Our model can be extended by incorporating additional information to the *DR*, such as some chronological information (e.g. number of turns up to the current turn) or user profiles (e.g. nave or experimented users or user preferences).

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.

As stated before, the *DR* contains information about concepts and attributes provided by the user throughout the previous history of the dialog. For the dialog manager to determine the next answer, we have assumed that the exact values of the attributes are not significant. They are important for accessing the databases and for constructing the output sentences of the system. However, the only information necessary to predict the next action by the system is the presence or absence of concepts and attributes. Therefore, the information we used from the *DR* is a codification of this data in terms of three values, $\{0, 1, 2\}$, for each field in the *DR* according to the following criteria:

- **0:** The concept is unknown or the value of the attribute is not given.
- **1:** The concept or attribute is known with a confidence score that is higher than a given threshold. Confidence scores are given during the recognition and understanding processes. and can be increased by means of confirmation turns.
- **2:** The concept or attribute is activated with a confidence score that is lower than the given threshold.

As an attempt to improve the current technology, we propose to merge statistical approaches with VoiceXML. Our goal is to combine the flexibility of statistical dialog management with the facilities that VoiceXML offers, which would help to introduce statistical approaches for the development of commercial (and not strictly

academic) conversational agents. To this end, our proposal employs the described statistical dialog management technique to decide the next system prompt. In addition, the system prompts and the grammars for ASR are implemented in VoiceXML compliant formats, for example, JSGF or SRGS.

In contrast with other hybrid approaches, our main aim is not to incorporate knowledge about best strategies in statistical dialog management, but rather to take advantage of an implementation language which has been traditionally used to build rule-based systems (such as VoiceXML), for the development of statistical dialog strategies. Expert knowledge about deployment of VoiceXML applications, development environments and tools can still be exploited using our technique. The only change is in the transition between states, which is carried out on a data-driven basis (i.e., is not deterministic). Figure 1 shows the complete architecture designed for the integration of the statistical methodology for dialog management with the functionalities provided by the VoiceXML standard.

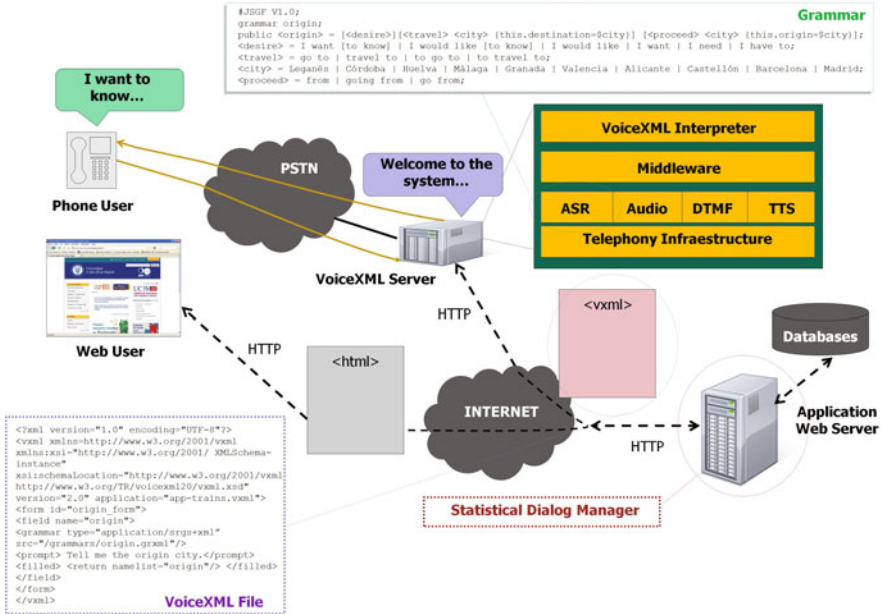


Fig. 1 Architecture for the integration of the proposed statistical methodology for dialog management with the VoiceXML standard

To learn the dialog model we use dialog simulation techniques. Our approach for acquiring a dialog corpus is based on the interaction of a user simulator and a dialog manager simulator [7]. 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.

The user simulation simulates the user intention level, that is, the simulator provides concepts and attributes that represent the intention of the user utterance. An error simulator module has been designed to perform error generation. In addition, the error simulator adds a confidence score to each concept and attribute in the frames.

A maximum number of turns per dialog is defined for acquiring a corpus using our user simulator, taking in to account the requirements of the task for real users. A user request for closing the dialog is selected once the system has provided the information defined in the objective(s) of the dialog. The dialogs that fulfill this condition before the maximum number of turns are considered successful. The dialog manager considers that the dialog is unsuccessful and decides to abort it when the following conditions take place: i) The dialog exceeds the maximum number of user turns; ii) The answer selected by the dialog manager corresponds with a query not required by the user simulator; iii) The database query module provides an error warning because the user simulator has not provided the mandatory information needed to carry out the query; iv) The answer generator provides a error warning when the selected answer involves the use of a data not provided by the user simulator.

3 Development of an Academic Conversational Agent

To test our proposal, we have used the UAH Task [4]. UAH is a spoken dialog system developed to provide spoken access to academic information about the Department of Languages and Computer Systems in the University of Granada. The information that the system provides can be classified in four main groups: subjects, professors, doctoral studies, and registration.

As in many other conversational agents, the semantic representation that we have chosen for the UAH task is based on the concept of frame. 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, Registration*), three task-independent concepts (*Affirmation, Negation, and Not-Understood*), and eight attributes (*Subject-Name, Degree, Group-Name, Subject-Type, Lecturer-Name, 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-GroupName*, *Confirmation-SubjectType*, *Confirmation-LecturerName*, *Confirmation-ProgramName*, *Confirmation-Semester*, and *Confirmation-Deadline*).

The DR defined for the UAH 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. The semantics selected for the dialog manager is represented through the 27 possible system answers defined for the UAH task. A set of 100,000 successfully simulated dialogs were used to learn the statistical dialog model.

A total of 25 dialogs were recorded from interactions of six students and professors of our University employing the conversational agent developed for the UAH task following our proposal. We considered the following measures for the evaluation:

1. Dialog success rate (%success). This is the percentage of successfully completed tasks. In each scenario, the user has to obtain one or several items of information, and the dialog success depends on whether the system provides correct data (according to the aims of the scenario) or incorrect data to the user.
2. Average number of turns per dialog (nT).
3. Confirmation rate (%confirm). It was computed as the ratio between the number of explicit confirmations turns (nCT) and the number of turns in the dialog (nCT/nT).
4. Average number of corrected errors per dialog (nCE). This is the average of errors detected and corrected by the dialog manager. We have considered only those errors that modify the values of the attributes and that could cause the failure of the dialog.
5. Average number of uncorrected errors per dialog (nNCE). This is the average of errors not corrected by the dialog manager. Again, only errors that modify the values of the attributes are considered.
6. Error correction rate (%ECR). The percentage of corrected errors, computed as $nCE / (nCE + nNCE)$.

The results presented in Table 1 show that the developed conversational can interact correctly with the users in most cases, achieving a success rate of 94%. The dialog success depends on whether the system provides the correct data for every objective defined in the scenario. The analysis of the main problems detected in the acquired dialogs shows that, in some cases, the system did not detect that the user wanted to finish the dialog. A second problem was related to the introduction of data in the DR with a high confidence value due to errors generated by the automatic speech recognizer that were not detected by the dialog manager. However, the evaluation confirms a good operation of the approach since the information is correctly given to the user in the majority of cases, as it is also shown in the value of the error correction rate.

Table 1 Results of the evaluation of the academic conversational agent

	%success	nT	%confirm	%ECR	nCE	nNCE
Conversational Agent	94%	11.6	38%	93%	0.89	0.06

4 Conclusions

In this paper, we have described a technique for integrating statistical techniques for the development of interactive conversational agents using a well known standard like VoiceXML, and considering a dialog model that is automatically learned from a dialog corpus. Our proposal works on the benefits of statistical methods for dialog management and VoiceXML, respectively. The former provide an efficient means to exploring a wider range of dialog strategies, whereas the latter makes it possible to benefit from the advantages of using the different tools and platforms that are already available to simplify system development.

We have applied our technique to develop a conversational agent that provides academic information, and have shown that it enables creating automatically VoiceXML documents to interact with the user. As a future work, we plan to study ways for adapting the proposed dialog management technique to more complex domains. Additionally, we are interested in investigating the main challenges that might arise in using our proposal to develop commercial systems, and the possible ways for easing the adoption of our technique in industry.

Acknowledgments. Funded by projects CICYT TIN2008-06742-C02-02/TSI, CICYT TEC 2008-06732-C02-02/TEC, CAM CONTEXTS (S2009/TIC-1485), and DPS2008-07029-C02-02.

References

1. Bohus, D., Grau, S., Huggins-Daines, D., Keri, V., Krishna, G., Kumar, R., Raux, A., Tomko, S.: Conquest - an Open-Source Dialog System for Conferences. In: Proc. of Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics, Rochester, NY, USA, pp. 9–12 (2007)
2. Bohus, D., Rudnicky, A.: Larri: A language-based maintenance and repair assistant. In: Proc. of IDS 2002, Kloster Irsee, Germany (2002)
3. Brahnam, S.: Building Character for Artificial Conversational Agents: Ethos, Ethics, Believability, and Credibility. *PsychNology Journal* 7(1), 9–47 (2009)
4. Callejas, Z., López-Cózar, R.: Relations between de-facto criteria in the evaluation of a spoken dialog system. *Speech Communication* 50(89), 646–665 (2008)
5. Bailly, G., Raidt, S.: Gaze, conversational agents and face-to-face communication. *Speech Communication* 52(6), 598–612 (2010)
6. Glass, J., Flammia, G., Goodine, D., Phillips, M., Polifroni, J., Sakai, S., Seneff, S., Zue, V.: Multilingual spoken-language understanding in the MIT Voyager system. *Speech Communication* 17, 1–18 (1995)

7. Griol, D., Hurtado, L., Sanchis, E., Segarra, E.: Acquiring and Evaluating a Dialog Corpus through a Dialog Simulation Technique. In: Proc. of the 8th SIGdial Workshop on Discourse and Dialogue, Antwerp, Belgium, pp. 39–42 (2007)
8. 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)
9. He, Y., Young, S.: A data-driven spoken language understanding system. In: Proc. of ASRU 2003, pp. 583–588. St. Thomas, U.S. Virgin Islands (2003)
10. López-Cozar, R., Araki, M.: Spoken, Multilingual and Multimodal Dialogue Systems. In: Development and Assessment. John Wiley and Sons, Chichester (2005)
11. Mazuel, L., Sabouret, N.: Generic command interpretation algorithms for conversational agents. *Web Intelligence and Agent Systems* 6(1), 43–57 (2008)
12. Melin, H., Sandell, A., Ihse, M.: Ctt-bank: A speech controlled telephone banking system - an initial evaluation. In: TMH-QPSR, vol. 1, pp. 1–27 (2001)
13. Menezes, P., Lerasle, F., Dias, J., Germa, T.: Towards an interactive humanoid companion with visual tracking modalities. *International Journal of Advanced Robotic Systems*, 48–78 (2007)
14. Minker, W., Waibel, A., Mariani, J.: Stochastically-Based Semantic Analysis. Kluwer Academic Publishers, Dordrecht (1999)
15. Rojas-Barahona, L., Giorgino, T.: Adaptable dialog architecture and runtime engine (adarte): A framework for rapid prototyping of health dialog systems. *International Journal of Medical Informatics* 78, 56–68 (2009)
16. Torres, F., Hurtado, L., García, F., Sanchis, E., Segarra, E.: Error handling in a stochastic dialog system through confidence measures. *Speech Communication* 45, 211–229 (2005)
17. Vaquero, C., Saz, O., Lleida, E., Marcos, J., Canalís, C.: VOCALIZA: An application for computer-aided speech therapy in spanish language. In: Proc. of IV Jornadas en Tecnología del Habla, Zaragoza, Spain, pp. 321–326 (2006)
18. Weng, F., Varges, S., Raghunathan, B., Ratiu, F., Pon-Barry, H., Lathrop, B., Zhang, Q., Scheideck, T., Bratt, H., Xu, K., Purver, M., Mishra, R., Raya, M., Peters, S., Meng, Y., Cavedon, L., Shriberg, L.: CHAT: A Conversational Helper for Automotive Tasks. In: Proc. of the 9th International Conference on Spoken Language Processing (Interspeech-ICSLP), Pittsburgh, USA, pp. 1061–1064 (2006)
19. Williams, J., Young, S.: Partially Observable Markov Decision Processes for Spoken Dialog Systems. *Computer Speech and Language* 21(2), 393–422 (2007)
20. Young, S.: The Statistical Approach to the Design of Spoken Dialogue Systems. Tech. rep., Cambridge University Engineering Department (2002)
21. Zue, V., Seneff, S., Glass, J., Polifroni, J., Pao, C., Hazen, T., Hetherington, L.: JUPITER: A telephone-based conversational interface for weather information. *IEEE Transactions on Speech and Audio Processing* 8, 85–96 (2000)