The Acquisition of a Dialog Corpus with a Prototype and two WOz[*]

L.F. Hurtado, E. Segarra, E. Sanchis, F. García
Dept. de Sistemes Informàtics i Computació
Universitat Politècnica de València
Camí de Vera sn, 46022 València
{lhurtado, esegarra, esanchis, fgarcia}@dsic.upv.es

D. Griol
Departamento de Informática
Universidad Carlos III de Madrid
Av. Universidad 30, 28911 Leganés
dgriol@inf.uc3m.es

Abstract

In this paper, we present our approach to simplify the dialog corpus acquisition task. This approach is based on the use of a prototype of the dialog manager and two Wizards of Oz.

1 Introduction

The development of spoken dialog systems is a complex process that involves the design, implementation and evaluation of a set of modules that deal with different knowledge sources. Currently, one of the most successful approaches is based on statistical models, which represent the probabilistic processes involved in each module, whose corresponding models are estimated by means of corpora of human-machine dialogs (Williams and Young, 2007; Griol et al., 2008). The success of statistical approaches highly depends on the quality of such models and, therefore, on the quality and size of the corpora from which they are trained. That is the reason why the acquisition of adequate corpora is a key process.

With the objective of facilitating the acquisition of a dialog corpus for the EDECAN-SPORT task for the booking of sports facilities within the framework of the EDECAN project (Lleida et al., 2006), we followed the process described below. Firstly, we analyzed human-human dialogs provided by the sports area of our university, which have the same domain defined for the EDECAN-SPORT task. From these dialogs we defined the semantics of the task in terms of dialog acts for both the user utterances and system prompts, and labeled these dialogs. Thus, we have a very low initial corpus for the EDECAN-SPORT task. From this small corpus we learned a preliminary version of the dialog manager (Griol et al.,

2008). This dialog manager was used as a prototype in the supervised process of acquiring a larger corpus by means of the Wizard of Oz technique.

Secondly, as the initial corpus is not large enough to train a suitable model for the speech understanding module, we do not have a preliminary version of this module for the acquisition process with the Wizard of Oz. Our proposal is based on using a specific Wizard of Oz to play the role of the natural language understanding module and a second Wizard of Oz to supervise the dialog manager. Using these two WOz allows us to obtain after the acquisition process not only the dialog corpus, but also the dialog acts corresponding to the labeling of the user and system turns (avoiding the subsequent process of manual labeling).

2 Architecture of the acquisition

Following the main contributions in the literature in the area of spoken dialog systems, we used the Wizard of Oz technique to acquire a dialog corpus for the EDECAN-SPORT task. The main difference of our proposal (Garcia et al., 2007) consists of using two Wizards of Oz: a simulator of the speech understanding process and a supervisor of the dialog manager. The first wizard listens to the user utterances, simulates the behavior of the automatic speech recognition and speech understanding modules for recognizing and understanding speech, and provides a semantic representation of the user utterance. From that representation, the second wizard supervises the behavior of the dialog manager. Figure 1 shows the architecture defined for the acquisition.

2.1 The understanding simulator

The Wizard of Oz that deals with the understanding process generates the semantic representation of the user utterances. To achieve the most similar result to a real statistical understanding module, the representation generated by the first wizard is
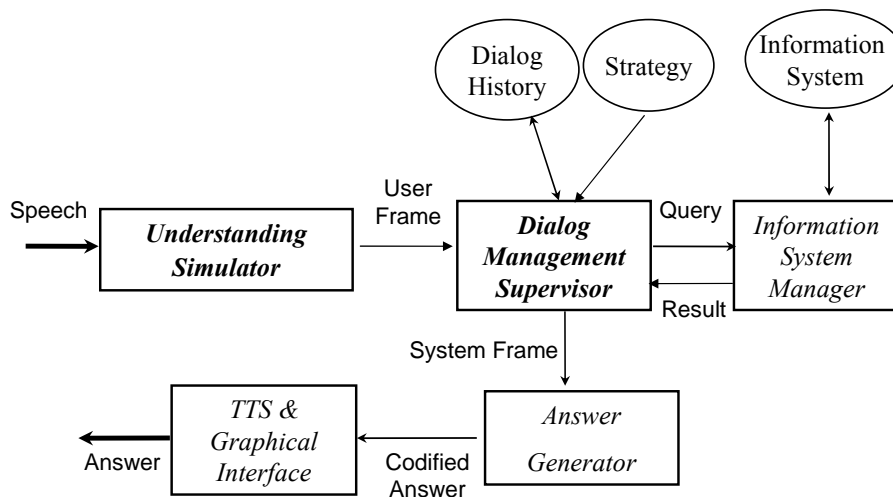
Figure 1: The proposed acquisition schema for the EDECAN corpus

passed though a module that simulates errors. This simulation (Garcia et al., 2007) is based on the analysis of the errors in the recognition and understanding processes generated when our models were applied to a corpus of similar characteristics.

2.2 The dialog manager

We have developed an approach to dialog management using a statistical model that is estimated from a dialog corpus (Griol et al., 2008). This model is automatically learned from a dialog corpus labeled in terms of dialog acts. This approach, which was originally developed to be used in a dialog system that provides train schedules and prices, was adapted for its use in the task of the booking sports facilities. This adaptation takes into account the new requirements introduced in this task, which involves using an application manager that interacts with the information servers and verifies if the user queries fulfill the regulations defined for the booking service. The actions taken by the application manager can affect the decision made by the dialog manager, aspect which was not considered in the previous task.

From the human-human corpus, a prototype of the dialog manager module was implemented to be included in our acquisition system. The second wizard supervised its behavior. This supervision is carried out by means of two applications. The first one is used to supervise the response automatically generated by the dialog manager (the wizard corrects this response when he considers that it is inadequate). The second application is used to supervise the operation of the application manager.

3 The acquisition

Using the approach described in this article, a set of 240 dialogs has been acquired for our task. A total of 18 different speakers from different origins (the headquarters of the research teams of the EDECAN consortium). The languages involved in the acquisition have been Spanish, Catalan and Basque. A set of 15 types of scenarios was defined in order to cover all the possible use cases of the task.

The information available for each dialog consists of four audio channels, the transcription of the user utterances (with an average of 5.1 user turns per dialog and 6.7 words per user turn) and the semantic labeling of the user and system turns.

References

F. Garcia, L.F. Hurtado, D. Griol, M. Castro, E. Segarra, and E. Sanchis. 2007. Recognition and Understanding Simulation for a Spoken Dialog Corpus Acquisition. In *TSD 2007*, volume 4629 of *LNAI*, pages 574–581. Springer.

D. Griol, L. F. Hurtado, E. Segarra, and E. Sanchis. 2008. A statistical approach to spoken dialog systems design and evaluation. In *Speech Communication*, volume 50, pages 666–682.

E. Lleida, E. Segarra, M. I. Torres, and J. Macías-Guarasa. 2006. EDECN: sistEma de Dilogo multidominio con adaptacin al contExto aCstico y de AplicaciN. In *IV Jornadas en Tecnologia del Habla*, pages 291–296, Zaragoza, Spain.

J. Williams and S. Young. 2007. Partially Observable Markov Decision Processes for Spoken Dialog Systems. In *Computer Speech and Language 21(2)*, pages 393–422.