Available online at www.sciencedirect.com
ScienceDirect
Computer Speech and Language xxx (2013) xxx–xxx
COMPUTER SPEECH AND LANGUAGE
www.elsevier.com/locate/csl

A domain-independent statistical methodology for dialog management in spoken dialog systems ☆

David Griol [a,*], Zoraida Callejas [b], Ramón López-Cózar [b], Giuseppe Riccardi [c]

[a] *Computer Science Department, Carlos III University of Madrid, Spain*
[b] *Department of Languages and Computer Systems, University of Granada, Spain*
[c] *Department of Information Engineering and Computer Science, University of Trento, Italy*

Received 3 October 2012; received in revised form 19 August 2013; accepted 8 September 2013

Abstract

This paper proposes a domain-independent statistical methodology to develop dialog managers for spoken dialog systems. Our methodology employs a data-driven classification procedure to generate abstract representations of system turns taking into account the previous history of the dialog. A statistical framework is also introduced for the development and evaluation of dialog systems created using the methodology, which is based on a dialog simulation technique. The benefits and flexibility of the proposed methodology have been validated by developing statistical dialog managers for four spoken dialog systems of different complexity, designed for different languages (English, Italian, and Spanish) and application domains (from transactional to problem-solving tasks). The evaluation results show that the proposed methodology allows rapid development of new dialog managers as well as to explore new dialog strategies, which permit developing new enhanced versions of already existing systems.

Keywords: Spoken dialog systems; Dialog management; Statistical methodologies; User modeling; Dialog simulation; Systems evaluation

1. Introduction

Spoken dialog systems are computer programs that receive as input speech and generate as output synthesized speech, engaging the user in a dialog that aims to be similar to that between humans (Pieraccini, 2012; Heinroth and Minker, 2012; López-Cózar and Araki, 2005; McTear, 2004; Gibbon et al., 2000). Thus, these interfaces make technologies more usable, as they ease interaction (Hempel, 2008), allow integration in different environments (Heinroth and Minker, 2012; Minker et al., 2004), and make technologies more accessible, especially for disabled people (Vipperla et al., 2012; Beskow et al., 2009).

Usually, SDSs carry out five main tasks: Automatic Speech Recognition (ASR), Spoken Language Understanding (SLU), Dialog Management (DM), Natural Language Generation (NLG), and Text-To-Speech Synthesis (TTS). These tasks are typically implemented in different modules of the system's architecture.

☆ This paper has been recommended for acceptance by A. Potamianos.

* Corresponding author. Tel.: +34 916249116.
E-mail addresses: david.griol@uc3m.es, dgriol@inf.uc3m.es (D. Griol), zoraida@ugr.es (Z. Callejas), rlopezc@ugr.es (R. López-Cózar), riccardi@disi.unitn.it (G. Riccardi).

0885-2308/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.csl.2013.09.002

The goal of speech recognition is to obtain the sequence of words uttered by a speaker (Tsilfidis et al., 2013; O'Shaughnessy, 2008; López-Cózar and Callejas, 2008). It is a very complex task, as there can be a great deal of variation in the input the recognizer must analyze, for example, in terms of the linguistics of the utterance, inter and intra speaker variation, the interaction context and the transmission channel. Once the speech recognizer has provided an output, the system must understand what the user said. The goal of spoken language understanding is to obtain the semantics from the recognized sentence. This process generally requires morphological, lexical, syntactical, semantic, discourse and pragmatical knowledge (Wu et al., 2010; López-Cózar et al., 2010; Minker, 1999).

The dialog manager decides the next action of the system (Traum and Larsson, 2003; Williams and Young, 2007; Griol et al., 2008), interpreting the incoming semantic representation of the user input in the context of the dialog. In addition, it resolves ellipsis and anaphora, evaluates the relevance and completeness of user requests, identifies and recovers from recognition and understanding errors, retrieves information from data repositories, and decides about the next system's response. Natural language generation is the process of obtaining sentences in natural language from the non-linguistic, internal representation of information handled by the dialog system (Lemon, 2011; López et al., 2011). Finally, the TTS module transforms the generated sentences into synthesized speech (Dutoit, 1996).

In order to enable rapid deployment of these systems, markup languages such as VoiceXML[1] have been widely adopted as they reduce the time and effort required for system implementation. However, system development with this approach involves a very costly engineering cycle (Rojas-Barahona and Giorgino, 2009). As an attempt to reduce this cost and carry out rapid system prototyping, statistical approaches are gaining increasing interest. These approaches enable automatic learning of dialog strategies, thus avoiding the time-consuming process that hand-crafted dialog design involves. Statistical models can be trained from real dialogs, modeling the variability in user behaviors. Although the construction and parameterization of these models depend on expert knowledge about the task to be carried out by the dialog system, the final objective is to develop systems that are more robust for real-world conditions, and that are easier to adapt to different users and tasks (Schatzmann et al., 2006).

In this paper we present a statistical approach for the development of dialog managers, which 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. Unlike other statistical approaches, our approach has the advantage of taking into account the data supplied by the user throughout the dialog, which makes the estimation of the statistical model tractable, without causing scalability problems. Our proposal is suitable to develop dialog managers regardless of the application domain and the interaction language. Moreover, it can be incrementally optimized to tackle complex tasks, as will be shown in the experiments.

After this introduction, the remainder of the paper is organized as follows. Section 2 describes existing approaches for the development of dialog managers, paying special attention to statistical approaches. Section 3 describes the proposed methodology for dialog management and the dialog generation technique employed to simulate dialogs. Section 4 describes the process and practical dialog systems used for evaluating our proposal. Section 5 presents the results of the evaluation of the dialog management methodology for the different systems using the proposed set of evaluation measures. Section 6 presents the conclusions and suggests some future work guidelines.

2. Related work

Although dialog management is only a part of the development cycle of spoken dialog systems, it can be considered one of the most demanding tasks given that this module encapsulates the logic of the speech application (Wilks et al., 2011). Traum and Larsson (2003) state that dialog management involves four main tasks: (i) updating the dialog context, (ii) providing a context for sentence interpretation, (iii) coordinating other modules and (iv) deciding the information to convey to the user and when to do it. Thus, the selection of a specific system action depends on multiple factors, such as the output of the speech recognizer (e.g., measures that define the reliability of the recognized information), the dialog interaction and previous dialog history (e.g., the number of repairs carried out so far), the application domain (e.g., guidelines for customer service), knowledge about the users, and the responses and status of external back-ends, devices, and data repositories. Given that the actions of the system directly impact users, the dialog manager is largely

[1] http://www.w3.org/TR/voicexml20/.

Please cite this article in press as: Griol, D., et al., A domain-independent statistical methodology for dialog management in spoken dialog systems. Comput. Speech Lang. (2013), http://dx.doi.org/10.1016/j.csl.2013.09.002