Employing Context of Use in Dialogue Processing

Botond Pakucs
Centre for Speech Technology (CTT)
KTH, Royal Institute of Technology
Lindstedtsvägen 24, 100 44 Stockholm, Sweden
botte@speech.kth.se

Abstract

In this paper, a generic solution is presented for capturing, representing and employing the context of use in dialogue processing. The implementation of the solution within the framework of the SesaME dialogue manager and the Butler demonstrator is also described.

1 Introduction

In natural human-to-human communication, speakers are able to use implicit contextual information to increase the conversational bandwidth. The implicit contextual information is relevant knowledge about the actual situation. However, this knowledge is not necessarily part of the linguistic context (has not been uttered earlier).

The following real life example illustrates the use of implicit knowledge about the situation.

An employee leaving the office:
Q: When is my train leaving?
A: At 17:30.

This dialogue appears somewhat strange, incomplete and even incomprehensible to others than the participants. In spite of this, the dialogue exhibits a successful interaction. This short dialogue appears to be a repetition of similar interactions encountered previously. The answer contains prediction and beliefs about the dialogue partner's individual goals, intentions and preferences.

For achieving more natural and efficient communication, it is desirable to enable spoken dialogue systems to support similar interactions.

2 Employing context of use in SesaME

SesaME (Pakucs, 2003) is a generic dialogue manager specially developed to enable multi-domain dialogues in mobile environments. The central idea is to know as much as possible about the users. Each user is expected to use an individual and *highly personalized speech interface* to access a multitude of services and appliances. This is achieved through a personalized speech interface integrated into some personal and wearable appliance such as a mobile phone or a PDA. The application specific data, including the dialogue management capabilities, is locally stored at the service provider side and is dynamically plugged into the personalized speech interface and activated whenever the user enters a new environment.

The SesaME architecture is comparable to other agent-based architectures such as the TRIPS (Allen et al., 2000) architecture. A central information storage, a blackboard, stores the representation of the system's *information state* (Larsson and Traum, 2000). However, this representation is not formalized; the information state is merely a collection of all data available to the dialogue system. An event-based solution is used for updating the information state, where events can be dialogue moves, internal system events, or changes in the user's context of use.

2.1 Knowledge representation

After each interaction with a user, every utterance and the related contextual information is represented as a feature vector containing feature-value pairs of all relevant information related to the in-

teraction. The only common property of the features in the feature vector is the co-occurrence. The feature vectors are indexed and stored in individual user models implemented as vector-space models. For manipulating the user models, well-known information retrieval solutions are used.

Albeit, every feature vector is domain and task dependent, the individual user models are generic and they may contain feature vectors from several different domains and tasks. In this way, capturing and employing cross-domain user characteristics is also feasible.

2.2 Context based adaptation

In SesaME, a content-based solution (Zukerman and Albrecht, 2001) is employed for performing a context-based adaptation to individual users.

A context manager keeps track of the user's current context. During every new interaction, based on the available contextual information one or more feature vectors are built. These vectors are used for retrieving similar interactions from the user model.

The retrieved results are used to predict specific features of the ongoing interaction and to achieve adaptation to the current context. For example, based on earlier interactions with a voice controlled travel-planer it may be possible to detect that a commuter's most frequent choice of destination on weekday evenings is "Stockholm". Thus, it is possible to ask the user a more natural question: *"Would you like a ticket to Stockholm?"* instead of the impersonal default prompt: *"Where would you like to travel?"*.

However, if no similar interactions are present in the user model, or no obvious patterns are detected, the default prompt is used.

3 Application and Evaluation

The Butler (Pakucs, 2004) is a telephony-based multi-domain dialogue system developed for evaluating the employment of contextual information in SesaME. The services provided by Butler can be categorized in three main categories, *public services* such as accessing commuter and subway train timetables, menu information for the nearby restaurants, *accessing personal information* from calendars and *accessing workplace related information*, such as time and location of meetings and seminars.

All these services are based on information services available on the Internet. The relevant knowledge is automatically extracted from the available html-documents and transformed into dialogue descriptions at runtime. The users are identified through speaker verification or based on the used mobile-phone numbers (A-numbers).

Currently a long-term evaluation of the Butler is conducted. However, preliminary data indicates that erroneous system predictions are considered as natural and non-disturbing by the users.

4 Conclusions

In this paper, a generic solution for employing implicit contextual knowledge in dialogue management was introduced. By encoding contextual information in the user model it becomes feasible to predict specific features of an ongoing interaction. Thus, a simultaneous adaptation to an individual user and the user's current situation is supported. This solution appears to be a promising contribution to providing a more flexible and natural interaction in spoken dialogue systems.

Acknowledgments

This research was carried out at the CTT, a competence center at KTH and was also sponsored by the European Union's IST Programme under contract IST-2000-29452, DUMAS.

References

James Allen, Donna Byron, Myroslava Dzikovska, George Ferguson, Lucian Galescu, and Amanda Stent. 2000. An architecture for a generic dialogue shell. *Natural Language Engineering*, 6(3-4):213–228, December.

Staffan Larsson and David R. Traum. 2000. Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural Language Engineering*, 6:323–340, September.

Botond Pakucs. 2003. Towards Dynamic Multi-Domain Dialogue Processing. In *Proceedings of the Eurospeech'03*, volume 1, pages 741–744, Geneva, Switzerland, September.

Botond Pakucs. 2004. Butler: A Universal Speech Interface for Mobile Environments. In *Proceedings of the Mobile HCI 04*, Glasgow, Scotland, September.

Ingrid Zukerman and David W. Albrecht. 2001. Predictive statistical models for user modeling. *User Modeling and User-Adapted Interaction*, 11:5–18.