Future Directions in Spoken Dialog Systems: A Community of Possibilities

Alan W Black and Maxine Eskenazi
Language Technologies Institute,
Carnegie Mellon University, Pittsburgh, PA, USA
{awb,max}@cs.cmu.edu

Abstract

A spoken dialog system consists of a number of non-trivially interacting components. In order to allow new students, researchers and developers to meaningfully and relatively rapidly enter the field it is critical that, despite their complexity, the resources be accessible and easy to use. Everyone should be able to start building new technologies without spending a significant amount of time re-inventing the wheel. There are four levels of support that we believe new entrants should have. 1) A flexible open source system that runs on many different operating systems, is well documented and supports both simple and complex dialog systems. 2) Logs and speech files from a large number of dialogs that enable analysis and training of new systems and techniques. 3) An actual set of real users that speak to the system on a regular basis. 4) The ability to run studies on complete real user platforms.

1 Background

The goal of the Dialog Research Center (DialRC) has been to provide the spoken dialog community with three levels of support in the form of tools and data for spoken dialog systems: open source software; logs and speech data from real dialogs, a community of real users that use a system regularly on real, useful platforms on which researchers can run studies. In this short paper we describe these four elements that our Center has endeavored to provide. Looking to the future we look to the spoken dialog community to contribute other platforms to ours to give newcomers to the field a rich set of experimental platforms on which to learn the ropes.

Open source spoken dialog software: We already provide, as Open Source software, the CMU Olympus Spoken Dialog System that offers ASR, TTS, a Dialog Manager and other components that allow developers to build both simple and complex dialog systems. While this architecture has been used in many systems (some of them are: TeamTalk [Harris et al 2004], RavenCalendar [Stenchikova et al. 2007], ConQuest [Bohus et al. 2007] and Let's Go [Raux et al. 2006]), it needs to be accompanied by more support in the form of both documentation and flexibility. It should also not be the only platform that is available to the community to run studies. There are new students, researchers and developers who want to hone their skills by adapting a dialog architecture and running it on a real user platform. In order to make it easier for these newcomers to build dialog systems in the form of short homework assignments (perhaps in 1-2 weeks), for a regular class, Olympus must be more flexible, and easier to understand and master.

With the open source core system that has already been released, we plan to add virtual machines that have all of the components preinstalled, as done in another area by [Tokuda et al 2012]. This will make it easier for newcomers to start writing and modifying dialog systems immediately rather than spending time installing black box software. This implies that our existing Windows support must be extended to also cover Linux and Mac OSX.

Log data from dialogs: Some of the significant, exciting advances that have recently been seen in the realm of spoken dialog systems use statistical modeling. This implies the acute need for data, above all *real data*, to train the models. The platforms that provide that data to a community should follow a standardized format in the same way that speech files have become standardized. Log Data can be used for offline analysis that, in turn, can

NAACL-HLT 2012 Workshop on Future directions and needs in the Spoken Dialog Community: Tools and Data, pages 19–20,
Montréal, Canada, June 7, 2012.

afford deeper first hand insight into how spoken dialog systems.

A community of real users and real platforms to run studies: we have seen [Young, 2010] that it is no longer reasonable to test a hypothesis about spoken dialog with a small number of paid participants. End users must be real: they have some interest in the outcome of the task at hand and they are not using the system just because they're paid and/or collecting evaluation data. This goal is difficult. However, we at DialRC want to provide a centralized mechanism to give the newcomers (and already established researchers as well) access to a group of platforms with real users. There must be the possibility of obtaining a tangible benefit from these real platforms and the research community should be willing to open their systems to others so that they can test their ideas in a realistic context, with a significant number of real callers.

Our current efforts have often centered on classic telephone-based information giving systems. Going forward it is important to take a wider view of the types of spoken dialog research we can address. Thus we are also interested in supporting: multimodal interaction, human-robot interaction, multi-party communication and even tasks with no clear definition of task completion (e.g. conversational banter). What is important is not promoting one type of research more than another. It is making many different real-user platforms available to the community at large.

CMU's DialRC proposes to act as a clearing house for software (our own and that of any others), data (both speech and logfiles), and runtime real application/real user platforms that gives the community a central place to find a platform that corresponds to their needs, to connect them to the developers of that platform, and to help distribute the data (speech and logfiles) coming from their use of it.

These three actions will build communities of new researchers and developers who, from their use of this plethora of platforms will enrich the latter with what they have learned and will enrich our community with their presence.

We envisage the following scenario. A student has a short assignment to make some change to one of the basic architectures that has been made available by DialRC. When the assignment is finished, they link their system to a platform that they found through our central listing. Real users call that platform (here, our student's system) when they need what is being offered (information on a good vegetarian restaurant in Cambridge, when the next bus to the airport is coming in Pittsburgh, a discussion of new things to see in a museum, etc). The student can then access the real user data that has been collected while their version of the system was running. Perhaps in a following assignment, if they provided two versions of the system, they can find out, from analyzing the data, which condition worked best. If they provided one condition, and other students provided other ones, then they can compare their results to those of the other students.

A constantly available resource, the competition between versions of a system does not have to be held at a time that may be inconvenient for some. It can be an ongoing event that researchers can participate in when it is convenient for them to do so.

References

Harris, T., Banerjee, S., Rudnicky, A., Sison, J., Bodine, K., and Black, A. (2004) "A Research Platform for Multi-Agent Dialogue Dynamics", IEEE Workshop on Robotics and Human Interactive Communication.

Stenchikova, S., Mucha, B., Hoffman, S., Stent, A. (2007), "RavenCalendar: A Multimodal Dialog System for Managing a Personal Calendar". HLT-NAACL 2007.

Bohus, B., Grau, S, Huggins-Danes, D., Keri, V., Anumachipalli, G., Kumar, R., Raux, A. & Tomko, S. (2007) "ConQuest: an Open-Source Dialog System for Conferences", HLT-NAACL 2007.

Raux, A., Langner, B., Bohus, D., Black, A., and Exkenazi, M. (2005) "Let's Go Public! Taking a Spoken Dialog System to the Real World." Interspeech 2005.

Tokuda, K, Lee, A, Yamagishi, J. et al. (2012) "Spoken Dialog system Framework based on User Generated Content", Nagoya Institute of Technology, and University of Edinburgh, funded under JST CREST Program.

Young, S. "Still Talking to Machines (Cognitively Speaking), Keynote Interspeech 2010, Makuhari Japan.