

6-2014

Knowledge Acquisition Strategies for Goal-Oriented Dialog Systems

Aasish K. Pappu
Carnegie Mellon University

Alexander I. Rudnicky
Carnegie Mellon University, ar28@andrew.cmu.edu

Follow this and additional works at: <http://repository.cmu.edu/lti>

 Part of the [Computer Sciences Commons](#)

Published In

Proceedings of the SIGDIAL Conference, 194-198.

This Conference Proceeding is brought to you for free and open access by the School of Computer Science at Research Showcase @ CMU. It has been accepted for inclusion in Language Technologies Institute by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

Knowledge Acquisition Strategies for Goal-Oriented Dialog Systems

Aasish Pappu Alexander I. Rudnicky

School of Computer Science

Carnegie Mellon University

{aasish, air}@cs.cmu.edu

Abstract

Many goal-oriented dialog agents are expected to identify slot-value pairs in a spoken query, then perform lookup in a knowledge base to complete the task. When the agent encounters unknown slot-values, it may ask the user to repeat or reformulate the query. But a robust agent can proactively seek new knowledge from a user, to help reduce subsequent task failures. In this paper, we propose knowledge acquisition strategies for a dialog agent and show their effectiveness. The acquired knowledge can be shown to subsequently contribute to task completion.

1 Introduction

Many spoken dialog agents are designed to perform specific tasks in a specified domain e.g., information about public events in a city. To carry out its task, an agent parses an input utterance, fills in slot-value pairs, then completes the task. Sometimes, information on these slot-value pairs may not be available in its knowledge base. In such cases, typically the agent categorizes utterances as non-understanding errors. Ideally the incident is recorded and the missing knowledge is incorporated into the system with a developer’s assistance — a slow offline process.

There are other sources of knowledge: automatically crawling the web, as done by NELL [Carlson et al., 2010], and community knowledge bases such as Freebase [Bollacker et al., 2008]. These approaches provide globally popular slot-values [Araki, 2012] and high-level semantic contexts [Pappu and Rudnicky, 2013]. Despite their size, these knowledge bases may not contain information about the entities in a specific target domain. However, users in the agent’s domain can potentially provide specific information on

slot/values that are unavailable on the web, e.g., regarding a recent interest/hobby of the user’s friend. Lasecki et al. [2013] have elicited natural language dialogs from humans to build NLU models for the agent and Bigham et al. [2010] have elicited answers to visual questions by integrating users into the system. One observation from this work is that both users and non-users can impart useful knowledge to system. In this paper we propose spoken language strategies that allow an agent to elicit new slot-value pairs from its own user population to extend its knowledge base. Open-domain knowledge may be elicited through text-based questionnaires from non-users of the system, but in a situated interaction scenario spoken strategies may be more effective. We address the following research questions:

1. *Can an agent elicit reliable knowledge about its domain from users?* Particularly knowledge it cannot locate elsewhere (e.g., on-line knowledge bases). Is the collective knowledge of the users sufficient to allow the agent to augment its knowledge through interactive means?
2. *What strategies elicit useful knowledge from users?* Based on previous work in common sense knowledge acquisition [Von Ahn, 2006, Singh et al., 2002, Witbrock et al., 2003], we devise spoken language strategies that allow the system to solicit information by presenting concrete situations and by asking user-centric questions.

We address these questions in the context of the `EVENTSPEAK` dialog system, an agent that provides information about seminars and talks in an academic environment. This paper is organized as follows. In Section 2, we discuss knowledge acquisition strategies. In Section 3, we describe a user study on these strategies. Then, we present an evaluation on system acquired knowledge and finally we make concluding remarks.

Table 1: System initiated strategies used by the agent for knowledge acquisition in the EVENTSPEAK system.

StrategyType	Strategy	Example Prompt
QUERYDRIVEN	QUERYEVENT QUERYPERSON	I know events on campus. What do you want to know? I know some of the researchers on campus. Whom do you want to know about?
PERSONAL	BUZZWORDS FAMOUSPEOPLE	What are some of the popular phrases in <i>your</i> research? Tell me some well-known people in <i>your</i> research area
SHOW&ASK	TWEET KEYWORDS PEOPLE	How would you describe this talk in a sentence, say a tweet. Give keywords for this talk in your own words. Do you know anyone who might be interested in this talk?

2 Knowledge Acquisition Strategies

We posit three different circumstances that can trigger knowledge acquisition behavior: (1) initiated by expert users of the system [Holzapfel et al., 2008, Spexard et al., 2006, Lütkebohle et al., 2009, Rudnicky et al., 2010], (2) triggered by “misunderstanding” of the user’s input [Chung et al., 2003, Filisko and Seneff, 2005, Prasad et al., 2012, Pappu et al., 2014], or (3) triggered by the system. They are described below:

QUERYDRIVEN. The system prompts a user with an open-ended question akin to “how-may-I-help-you” to learn what “values” of a slot are of interest to the user. This strategy does not ground user about system’s knowledge limitations. However, it allows the system to acquire information (slot-value pairs) from user’s input. The system can choose to respond to the input or ignore the input depending on its knowledge about the slot-value pairs in the input. Table 1 shows strategies of this kind i.e., QUERYEVENT and QUERYPERSON.

PERSONAL. The system asks a user about their own interests and people who may share those interests. This is an open-ended request as well, but the system expects the response to be confined to the user’s knowledge about specific entities in the environment. BUZZWORDS and FAMOUSPEOPLE expects the user to provide values for the slots.

SHOW&ASK. The system provides a description of an event and asks questions to ground user’s responses in relation to that event. E.g., given the title and abstract of a technical talk, the system asks the user questions about the talk. TWEET strategy is expected to elicit a concise description of the event, which eventually may help the agent to both summarize events for other users and identify keywords for an event. KEYWORDS strategy expects the user to explicitly supply keywords for an event. PEOPLE strategy expects the user to provide names of likely event participants.

We hypothesized that these strategies may allow the agent to learn new slot-value pairs that may

help towards better task performance.

3 Knowledge Acquisition Study

We conducted a user study to determine reliability of the information acquired by the system. We performed this study using the EVENTSPEAK¹ dialog system, which provides information about upcoming talks and other events that might be of interest, and about ongoing research on campus. The system presents material on a screen and accepts spoken input, in a context similar to a kiosk.

The study evaluated performance of the seven strategies described above. For SHOW&ASK strategies, we had users respond regarding a specific event. We used descriptions of research talks collected from the university’s website. We used a web-based interface for data collection; the interface presented the prompt material and recorded the subject’s voice response. Testvox² was used to setup the experiments and Wami³ for audio recording.

3.1 User Study Design

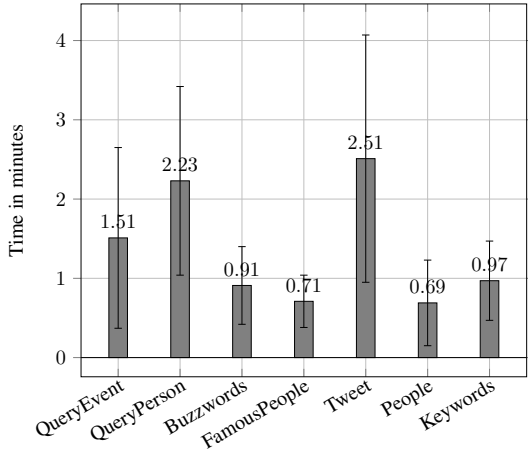
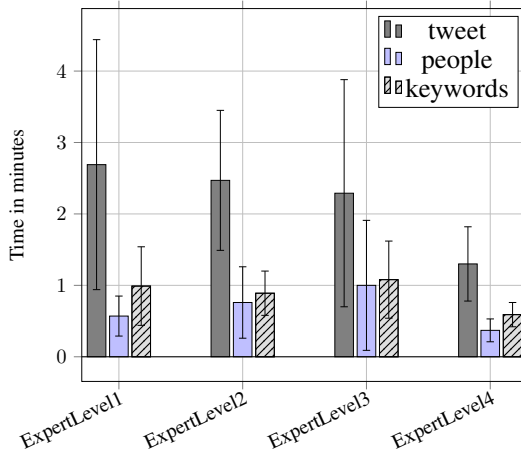
We recruited 40 researchers (graduate students) from the School of Computer Science, at Carnegie Mellon, representative of the user population for the EVENTSPEAK dialog system. Each subject responded to prompts from the QUERYDRIVEN, PERSONAL and SHOW&ASK strategies.

In the QUERYDRIVEN tasks, the QUERYEVENT strategy, the system responds to the user’s query with a list of talks. The user’s response is recorded, then sent to an open-vocabulary speech recognizer; the result is used as a query to a database of talks. The results are then displayed on the screen. The system applies the QUERYPERSON strategy in a similar way. In the PERSONAL tasks, the system applies the BUZZWORDS strategy to ask the user about popular keyphrases in their research

¹<http://www.speech.cs.cmu.edu/apappu/kacq>

²<https://bitbucket.org/happyalu/testvox/wiki/Home>

³<https://code.google.com/p/wami-recorder/>

Figure 1: Time per Task for all strategies**Figure 2: Time per Task vs Expertise**

area. The system then asks about well-known researchers (FAMOUSPEOPLE) in the user’s area.

In the SHOW&ASK tasks, we use two seminar descriptions per subject (in our pilot study, we found that people provide more diverse responses (in term of entities) in the SHOW&ASK based on the event abstract, compared to PERSONAL, QUERYDRIVEN). We used a set of 80 research talk announcements (consisting of a title, abstract and other information). For each talk, the system used all three strategies viz., TWEET, KEYWORDS and PEOPLE. For the TWEET tasks, subjects were asked to provide a one sentence description. They were allowed to give a non-technical/high-level description if they were unfamiliar with the topic. For the PEOPLE task, subjects had to give names of colleagues who might be interested in the talk. For the KEYWORDS task, subjects provided keywords, either their own words or ones selected from the abstract.

Since the material is highly technical, we were interested whether the tasks are cognitively demanding for people who are less familiar with the subject of a talk. Therefore, users were asked to indicate their familiarity with a particular talk (research area in general) using a scale of 1–4: 4 being more familiar and 1 being less familiar.

3.2 Corpus Description

This user study produced 64 minutes of audio data, on average 1.6 minutes per subject. We transcribed the speech then annotated the corpus for people names, and for research interests. Table 2 shows the number of unique slot-values found in the corpus. We observe that the number of unique research interests produced during SHOW&ASK is higher than for other strategies. This confirms

our initial observations that this strategy elicits diverse responses. The PERSONAL task produced a relatively higher number of researcher names (FAMOUSPEOPLE strategy) than other tasks. One explanation might be that people may find it easier to recall names in their own research area, as compared to other areas. Overall, we identified 139 unique researcher names and 485 interests.

Table 2: Corpus Statistics

StrategyType	Unique Researcher Names	Unique Research Interests
QUERYDRIVEN	21	30
PERSONAL	77	107
SHOW&ASK	76	390
Overall	139	485

3.3 Corpus Analysis

One of the objectives of this work is to determine *What strategies can the agent use to elicit knowledge from users?* Although, time-cost will vary with task and domain, a usable strategy should, in general, be less demanding. We analyzed the time-per-task for each strategy, shown in Figure 1. We found that the TWEET strategy is not only more demanding, it has higher variance than other tasks. One explanation is that people would attempt to summarize the entire abstract including technical details, despite the instructions indicated that a non-technical description was acceptable. We can see a similar trend in Figure 2 that irrespective of expertise-level, subjects take more time to give one sentence descriptions. We also observe high variance and higher time-per-task for QUERYPERSON; this is due to the system deliberately not returning any results for this task. This was done to

Table 3: Mean Precision for 200 researchers, broken down by the “source” strategy used to acquire their name
Note: Only 85 of 200 researchers had Google Scholar pages, GScholar Accuracy is computed for only those 85.

Metric	Description Text	SHOW&ASK	PERSONAL	QUERYDRIVEN	mean
Mean Precision	89.5%	86.9%	93.6%	86.2%	90.5%
GScholar Acc.	78.3%	82.3%	86.1%	100%	80.0%

find out whether subjects would repeat the task on failure. Ideally the system needs to only rarely use this strategy to not lose user’s trust and solicit multiple values for a given slot (e.g., person name) as opposed to requesting list of values as in FAMOUSPEOPLE and PEOPLE strategies. We find that PEOPLE, KEYWORDS, FAMOUSPEOPLE and BUZZWORDS strategies are efficient with a time-per-task of less than one minute. As shown in Figure 2, subjects do not take much time to speak a list of names or keywords.

4 Evaluation of Acquired Knowledge

To answer *Can an agent elicit reliable knowledge about its domain from users?* we analyzed the relevance of acquired knowledge. We have two disjoint list of entities, (a) researchers and (b) research interests; in addition we have speaker names from the talk descriptions. Our goal is to implicitly infer a list of interests for each researcher without soliciting the user for the interests of every researcher exhaustively. To each researcher in the list, we attribute list of interests that were mentioned in the same context as researcher was mentioned. We tag list of names acquired from the FAMOUSPEOPLE strategy with list of keywords acquired from the BUZZWORDS strategy — both lists acquired from same user. We repeat this process for each name mentioned in relation to a talk in the SHOW&ASK strategy. We tag keywords mentioned in the KEYWORDS strategy to researchers mentioned in the PEOPLE strategy.

4.1 Analysis

We produced 200 entries for researchers and their set of interests. We then had two annotators (senior graduate students) mark whether the system-predicted interests were relevant/accurate. The annotators were allowed to use information found on researchers’ home pages and Google Scholar⁴ to evaluate the system-predicted interests.

This can be seen as an information retrieval (IR) problem, where researcher is “query” and interests are “documents”. So, we use *Mean Precision*, a

common metric in IR, to evaluate retrieval. In our case, the ground truth for relevant interests comes from the annotators. The results are shown in Table 3. Our approach has high precision, 90.5%, for all 200 researchers. We see that irrespective of the strategy used to acquire entities, precision is good. We also compared our predicted interests with interests listed by researchers themselves on Google Scholar. There are only 85 researchers from our list with a Google Scholar page; for these our accuracy is 80%, again good. Moreover, significant knowledge is absent from the web (at least in our domain) yet can be elicited from users familiar with the domain.

5 Conclusion

We describe a set of knowledge acquisition strategies that allow a system to solicit novel information from users in a situated environment. To investigate the usability of these strategies, we conducted a user study in the domain of research talks. We analyzed a corpus of system-acquired knowledge and have made the material available⁵. Our data show that users on average take less than a minute to provide new information using the proposed elicitation strategies. The reliability of acquired knowledge in predicting relationships between researchers and interests is quite good, with a mean precision of 90.5%. We note that the PERSONAL strategy, which tries to tap personal knowledge, appears to be particularly effective. More generally, automated elicitation appears to be a promising technique for continuous learning in spoken dialog systems.

6 Appendix

—— System Predicted Researcher-Interests 1 ——
rich stern deep neural networks, speech recognition, signal processing, neural networks, machine learning, speech synthesis

⁴scholar.google.com

⁵www.speech.cs.cmu.edu/apappu/pubdl/eventspeak_corpus.zip

—— System Predicted Researcher-Interests 2 ——
kishore prahallad dialogue systems, prosody, speech synthesis, text to speech, pronunciation modeling, low resource languages

—— System Predicted Researcher-Interests 3 ——
carolyn rose crowdsourcing, meta discourse classification, statistical analysis, presentation skills instruction, man made system, education models, human learning

—— System Predicted Researcher-Interests 4 ——
florian metze dialogue systems, speech recognition, nlp, prosody, speech synthesis, text to speech, pronunciation modeling, low resource languages, automatic accent identification

—— System Predicted Researcher-Interests 5 ——
madhavi ganapathiraju protein structure, continuous graphical models, generative models, structural biology, protein structure dynamics, molecular dynamics

—— System Predicted Researcher-Interests 6 ——
alexander hauptmann discriminatively trained models, deep learning, computer vision, big data

—— System Predicted Researcher-Interests 7 ——
jamie callan learning to rank, search, large scale search, web search, click prediction, information retrieval, web mining, user activity, recommendation, relevance, machine learning, web crawling, distributed systems, structural similarity

—— System Predicted Researcher-Interests 8 ——
lori levin natural language understanding, knowledge reasoning, construction grammar, knowledge bases, natural language processing

References

- Masahiro Araki. Rapid development process of spoken dialogue systems using collaboratively constructed semantic resources. In *Proceedings of the SIGDIAL 2012 Conference*, pages 70–73. ACL, 2012.
- Jeffrey P Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samuel White, et al. Vizwiz: nearly real-time answers to visual questions. In *Proceedings of the 23rd ACM Symposium on User Interface software and technology*, pages 333–342. ACM, 2010.
- Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. *Proceedings of the SIGMOD*, pages 1247–1249, 2008.
- Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R Hruschka Jr., and Tom M Mitchell. Toward an Architecture for Never-Ending Language Learning. *Artificial Intelligence*, 2(4):1306–1313, 2010.
- Grace Chung, Stephanie Seneff, and Chao Wang. Automatic acquisition of names using speak and spell mode in spoken dialogue systems. In *Proceedings of the NAACL-HLT*, pages 32–39. ACL, 2003.
- Edward Filisko and Stephanie Seneff. Developing city name acquisition strategies in spoken dialogue systems via user simulation. In *6th SIGdial Workshop on Discourse and Dialogue*, 2005.
- Hartwig Holzapfel, Daniel Neubig, and Alex Waibel. A dialogue approach to learning object descriptions and semantic categories. *Robotics and Autonomous Systems*, 56(11): 1004–1013, November 2008.
- Walter Stephen Lasecki, Ece Kamar, and Dan Bohus. Conversations in the crowd: Collecting data for task-oriented dialog learning. In *First AAAI Conference on Human Computation and Crowdsourcing*, 2013.
- Ingo Lütkebohle, Julia Peltason, Lars Schillingmann, Christof Elbrechter, Britta Wrede, Sven Wachsmuth, and Robert Haschke. The Curious Robot: Structuring Interactive Robot Learning. In *ICRA’09*, pages 4156–4162. IEEE, 2009.
- Aasish Pappu and Alexander Rudnicky. Predicting tasks in goal-oriented spoken dialog systems using semantic knowledge bases. In *Proceedings of the SIGDIAL*, pages 242–250. ACL, 2013.
- Aasish Pappu, Teruhisa Misu, and Rakesh Gupta. Investigating critical speech recognition errors in spoken short messages. In *Proceedings of IWSDS*, pages 39–49, 2014.
- Rohit Prasad, Rohit Kumar, Sankaranarayanan Ananthakrishnan, Wei Chen, Sanjika Hewavitharana, Matthew Roy, Frederick Choi, Aaron Challenner, Enoch Kan, Arvind Neelakantan, et al. Active error detection and resolution for speech-to-speech translation. In *Proceedings of IWSLT*, 2012.
- Alexander I Rudnicky, Aasish Pappu, Peng Li, and Matthew Marge. Instruction Taking in the TeamTalk System. In *Proceedings of the AAAI Fall Symposium on Dialog with Robots*, pages 173–174, 2010.
- Push Singh, Thomas Lin, Erik T Mueller, Grace Lim, Travell Perkins, and Wan Li Zhu. Open mind common sense: Knowledge acquisition from the general public. In *CoopIS, DOA, and ODBASE*, pages 1223–1237. Springer, 2002.
- Thorsten Spexard, Shuyin Li, Britta Wrede, Jannik Fritsch, Gerhard Sagerer, Olaf Booij, Zoran Zivkovic, Bas Terwijn, and Ben Krose. BIRON, where are you? Enabling a robot to learn new places in a real home environment by integrating spoken dialog and visual localization. *Integration The VLSI Journal*, (section II):934–940, 2006.
- Luis Von Ahn. Games with a purpose. *Computer*, 39(6): 92–94, 2006.
- Michael Witbrock, David Baxter, Jon Curtis, Dave Schneider, Robert Kahlert, Pierluigi Miraglia, Peter Wagner, Kathy Panton, Gavin Matthews, and Amanda Vizedom. An interactive dialogue system for knowledge acquisition in cyc. In *Proceedings of the 18th IJCAI*, pages 138–145, 2003.