Probabilistic, Multi-staged Interpretation of Spoken Utterances

Ingrid Zukerman, Michael Niemann, Sarah George and Yuval Marom

Faculty of Information Technology, Monash University
Clayton, VICTORIA 3800, AUSTRALIA
{ingrid,niemann,sarahg,yuvalm}@csse.monash.edu.au

Abstract. We describe *Scusi?*, a multi-stage, spoken language interpretation mechanism designed to be part of a robot-mounted dialogue system. *Scusi?*'s interpretation process maps spoken utterances to conceptual graphs, and the nodes in these graphs to concepts in the world. Maximum posterior probability is used to rank the (partial) interpretations produced at each stage of this process.

1 Introduction

The DORIS project aims to develop a spoken dialogue module for a robotic agent. Eventually, this module will be able to engage in a dialogue with users and plan physical actions (by interfacing with a planner). In this paper, we describe *Scusi?*, the speech interpretation module that is being developed within the DORIS framework. *Scusi?* was designed to achieve the following objectives: (1) provide a variety of possible interpretations of the input, ranked according to merit; (2) change its mind regarding the best interpretation if warranted by new information; and (3) assist the dialogue module in handling partial or faulty interpretations and recovering from them.

To achieve these objectives we use a multi-stage, probabilistic interpretation process based on that introduced in [1]. Each interpretation is an understanding of spoken input in terms of the system's information structures and domain knowledge. Our interpretation process comprises three main stages: (a) speech recognition, (b) parsing, and (c) semantic interpretation (Section 2). Each stage produces multiple candidate options, which are ranked according to their probability of being intended by the speaker. The probability of a candidate depends on the probability of its parents (generated in the previous stage of the interpretation process) and that of its components (Section 4).

Each of the final interpretations and the intermediate *sub-interpretations* are maintained by *Scusi?*, enabling it to return to previously "unpreferred" options, update previous sub-interpretations in light of new information, and possibly change its mind regarding the preferred interpretation. Additionally, by keeping track of the probability of different components of an interpretation, *Scusi?* can determine which parts of an interpretation can be used for further processing, and which should be re-examined in order to get a better interpretation.

2 Multi-Stage Processing

Figure 1 illustrates the stages involved in processing spoken input. The first stage activates ViaVoice – an *Automatic Speech Recognizer (ASR)* – to generate candidate sequences of words (*Text*) from a *Speech Wave*. Each Text has a score that represents how

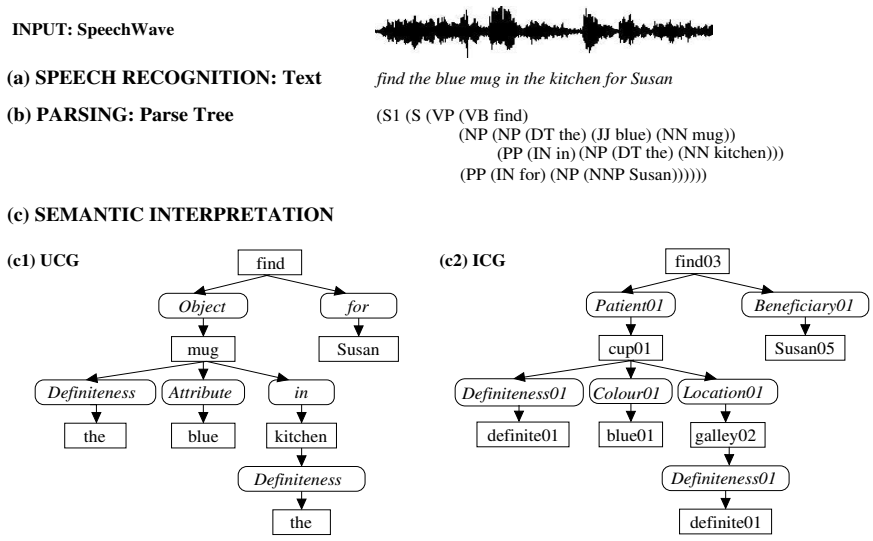

Fig. 1. Structures for the interpretation stages

well its words fit the speech wave. This score is converted into a probability. The word sequences are then parsed using Charniak's probabilistic parser, which generates a set of *Parse Trees*. The last stage uses *Concept Graphs (CGs)* to perform semantic interpretation. Two types of CGs are generated (Section 3). First, the Parse Trees are mapped into a set of *Uninstantiated Concept Graphs (UCGs)*, which represent mainly syntactic information. Next, *Scusi?* proposes candidate *Instantiated Concept Graphs (ICGs)* from selected UCGs. This is done by nominating *Instantiated Concepts* from DORIS's knowledge base as a potential realization for every *Uninstantiated Concept* in a UCG.

2.1 Anytime processing

The consideration of all possible options for each stage of the interpretation process is computationally intractable. To address this problem, we have adapted the *anytime* algorithm described in [1], which applies a selection-expansion cycle to build a search graph as follows. The selection step nominates a single sub-interpretation (Speech, Text, Parse Tree or UCG) to expand, and the expansion step generates only one child for that sub-interpretation. The selection step then nominates the next sub-interpretation to expand, which may be the one that was just expanded, its new child, or any other sub-interpretation in the search graph.

The nomination of a sub-interpretation is made on the basis of two factors: (1) the level in the search tree, and (2) the probability of its previously generated children. Preference is given to later stages in the search (e.g., UCGs rather than Texts) to encourage the early generation of complete interpretations, and within a stage, to sub-interpretations that have a "proven track record", i.e., that have previously produced high-probability children.

The selection-expansion cycle is repeated until one of the following happens: all the options are fully expanded, a time limit is reached, or the system runs out of memory. The results of the interpretation process are available to the dialogue module contin-

uously. This enables the dialogue module to decide on an action on the basis of the progress of the interpretation process.

2.2 Features of the interpretation process

Our interpretation process implements an *iterative optimization method*, which retains all sub-interpretations, and allows even low-ranking sub-interpretations to generate children. This safeguards against getting stuck in local maxima, and supports a flexible and dynamic behaviour, allowing DORIS to improve its understanding and management of the dialogue in various key ways.

- *Reviewing the overall picture:* Each sub-interpretation is stored with its individual probability. As a result, the dialogue module is not restricted to choosing its actions solely on the basis of the high ranking ICGs. It can use other information obtained during the interpretation process to determine whether special attention needs to be paid to the results from other stages in the interpretation process. For instance, if the Text for the current top ICG has a low-probability due to the speech recognizer "mishearing" an unclear speech wave, the dialogue module can initiate a clarification sub-dialogue regarding the original input.
- *Recovery from partial success:* Interpretations that do not match all grammatical sub-categorization requirements or domain expectations have a low probability, but are still retained. This enables the dialogue module to initiate a focused recovery using elements from incomplete or flawed interpretations.
- *Updating probabilities:* New information can be added to concept graphs (e.g., as a result of user feedback or visual input) or to the knowledge base (e.g., after an action performed by a planner in the physical world). Our process allows the probabilities of existing sub-interpretations and final interpretations (ICGs) to be efficiently recalculated and re-ranked in light of updates.
- *Background exploration:* Provided memory is available, *Scusi?* continues its search in the background, even after the time limit is reached. If a significantly better ICG is found, then a new interpretation can be proposed to the dialogue module.

3 Conceptual Graphs

Conceptual graphs represent entities and the relationships between them [2]. For instance, the CG in Figure 1(c2) indicates that there are two concepts find03 and cup01 that have a Patient01 Relationship.

Uninstantiated Conceptual Graphs. A UCG represents the syntactic relationships found in a Parse Tree. Most phrases map to a Concept Node representing their head-word. Slightly different Parse Trees may yield the same UCG. For instance, the blue Concept Node in Figure 1(c1) could be generated from an Adjectival Phrase instead of a stand-alone adjective (JJ) adjunct to the NP. The mapping from Parse Tree to UCG works the same way for declarative, imperative and interrogative sentences and for single words.

Instantiated Conceptual Graphs and the Knowledge Base. The generation of an ICG requires the selection of an *Instantiated Concept* from the knowledge base for each *Uninstantiated Concept* in a UCG. The knowledge base contains entries for specific real-world objects (e.g., cup03), general objects (e.g., CupClass01), domain actions (e.g., find02), abstract concepts (e.g., blue01) and relationships (e.g., Patient01). To

postulate Instantiated Concepts for Uninstantiated ones, each Uninstantiated Concept in a UCG is associated with a list of Instantiated Concepts. Each entry in the list is assigned a probability on the basis of how well it matches the Uninstantiated Concept. To generate an ICG, one Instantiated Concept is iteratively selected from the list of each Uninstantiated Concept in the parent UCG, starting with the most probable candidates.

4 Probabilities of Interpretations

ICGs are ranked according to their posterior probability in light of the given speech wave and the conversational context, which is obtained from concepts recently seen or mentioned. The interpretation stages shown in Figure 1 lead to the following equation for the probability of an ICG.

$$\Pr(ICG|Speech, Context) = \alpha \Pr(ICG|Context) \times \sum_{txt,prsTr,ucg} \left\{ \begin{matrix} \Pr(UCG|ICG) \times \Pr(ParseTr|UCG) \times \\ \Pr(Text|ParseTr) \times \Pr(Speech|Text) \end{matrix} \right\}$$

where α is a normalizing constant. The summation is required since a sub-interpretation may have multiple parents.

At present, the probabilities required to perform this calculation are not available. Instead of providing probabilities in the required direction, the ASR provides probabilities from Speech Wave to Text (its scores are directly translated to probabilities), and the probabilistic parser from Text to Parse Tree. Hence, we approximate the calculation of the posterior probability of an ICG as follows.

$$\Pr(ICG|Speech, Context) \cong \alpha \Pr(ICG|Context) \times \sum_{txt,prsTr,ucg} \left\{ \begin{matrix} \Pr(ICG|UCG) \times \Pr(UCG|ParseTr) \times \\ \Pr(ParseTr|Text) \times \Pr(Text|Speech) \end{matrix} \right\}$$

We assume that $\Pr(UCG|ParseTr)=1$ since a UCG is directly built from the Parse Tree. $\Pr(ICG|UCG)$ depends only on the match between each ICG concept and its corresponding UCG concept, given the lexical items associated with the ICG concept. $\Pr(ICG|Context)$ represents whether the concepts in the ICG make sense in the current context, given their salience and the relationships they expect to be in.

5 Related Research

This research extends our previous work [1] in its use of conceptual graphs as a representation formalism and its expansion of the probabilistic interpretation framework to incorporate these knowledge structures.

Conceptual graphs were also used by Sowa and Way [3] and Shankaranarayanan and Cyre [4] for discourse interpretation. However, these researchers did not employ a probabilistic approach, did not retain multiple interpretations, and their usage of conceptual graphs resembled semantic grammars, which take advantage of domain related information early in the interpretation process. Miller *et al.* [5] and He and Young [6] also used semantic grammars for the interpretation of utterances from the ATIS corpus, but they applied a probabilistic approach similar to ours. Instead of using semantic grammars, *Scusi?*'s interpretation process initially uses generic, syntactic tools, and incorporates semantic- and domain-related information only in the final stage.

Our work resembles most that of Gorniak and Roy [7] in its use of a probabilistic parser and its integration of context-based expectations with alternatives obtained from spoken utterances. But Gorniak and Roy do not retain multiple interpretations, and they restrict the search space by training the parser on a domain-relevant corpus and providing tightly constrained domain expectations.

6 Conclusion

We have presented a multi-stage, probabilistic interpretation process that derives the meaning of spoken utterances in light of contextual information. Our interpretation process relies on an approximation of Bayesian propagation for the incremental calculation of the probability of interpretations; an iterative optimization method, which allows the examination of suboptimal sub-interpretations; and a unified knowledge representation formalism that can represent information from various sources.

Additionally, our interpretation process brings to bear semantic and domain knowledge only in the final stage of this process, retains all sub-interpretations and information about them, and can take new information into account. These features enable our system to make some sense of flawed interpretations, change its mind about the best interpretation, and provide sufficient information to a dialogue module for handling partial or flawed interpretations. Preliminary tests indicate that improved results are gained by examining sub-interpretations that are not ranked highly.

Acknowledgments

This research was supported in part by the ARC Centre for Perceptive and Intelligent Machines in Complex Environments. The authors thank Eugene Charniak for his modifications to his probabilistic parser, and Charles Prosser for his assistance in extracting multiple texts from ViaVoice.

References

1. Niemann, M., George, S., Zukerman, I.: Towards a probabilistic, multi-layered spoken language interpretation system. In: Proceedings of the Fourth IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems, Edinburgh, Scotland (2005) 8–15
2. Sowa, J.: Conceptual Structures: Information Processing in Mind and Machine. Addison-Wesley (1984)
3. Sowa, J., Way, E.: Implementing a semantic interpreter using conceptual graphs. IBM Journal of Research and Development **30**(1) (1986) 57–69
4. Shankaranarayanan, S., Cyre, W.: Identification of coreferences with conceptual graphs. In: ICCS'94 – Proceedings of the Second International Conference on Conceptual Structures, College Park, Maryland (1994)
5. Miller, S., Stallard, D., Bobrow, R., Schwartz, R.: A fully statistical approach to natural language interfaces. In: ACL96 – Proceedings of the 34th Conference of the Association for Computational Linguistics, Santa Cruz, California (1996) 55–61
6. He, Y., Young, S.: A data-driven spoken language understanding system. In: ASRU'03 – Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop, St. Thomas, US Virgin Islands (2003)
7. Gorniak, P., Roy, D.: Probabilistic grounding of situated speech using plan recognition and reference resolution. In: ICMI'05 – Proceedings of the Seventh International Conference on Multimodal Interfaces, Trento, Italy (2005)