Sigma Spoken Language Understanding System

Thesis proposal

Himanshu Joshi, advised by Dr. Paul Rosenbloom, University of Southern California (USC) (http://cs.usc.edu/~rosenblo/)

ABSTRACT

This thesis proposes the integration of incremental speech processing with language understanding and cognition. Speech signal obtained from a typical speech front end shall be combined with linguistic knowledge in the form of phonetic, syntactic and semantic knowledge sources with cognition selecting the most likely word incrementally. This non-modular, supraarchitectural integration of spoken language has never been attempted in cognitive architectures making this work novel.

A cognitive architecture is a hypothesis about the fixed structures underlying intelligent behavior. Cognitive architectures support an important goal of AI – to understand and create synthetic agents with human capabilities (Langley, Laird, & Rogers, 2009). Integration across a wide range of capabilities is a key requirement for cognitive architectures (Rosenbloom, 2015) (Langley, Laird, & Rogers, 2009). Speech understanding is an important cognitive capability and yet not addressed by existing cognitive architectures – indicating the mixed (symbolic and probabilistic) and hybrid (discrete and continuous) nature of the speech problem. Traditional symbolic architectures such as Soar (Laird, 2012) interface to sub-cognitive modules outside of the core architecture for perceptual processing. Connectionist approaches (Sun, Merrill, & Peterson, 2001), (O'Reilly, 1998) (O'Reilly, Hazy, & Herd, 2012) do a good job of processing sub-symbolic input but do not have the symbolic capabilities to induce a breadth of intelligent capabilities (Langley, Laird, & Rogers, 2009). Even hybrid architectures such as CLARION (Sun, 2006) or SAL (Lebiere, O'Reilly, Jilk, Taatgen, & Anderson, 2008) have not tackled the speech problem.

Sigma is a new breed of cognitive architecture that aims to ultimately explain human cognition in terms of the capabilities it integrates and the interactions amongst them. In pursuit of this goal, Sigma is guided by four desiderata: (i) *grand unification*, aiming to integrate both symbolic and key sub-symbolic (perceptual) capabilities, (ii) *functional elegance*, aiming to derive cognitive and sub-cognitive capabilities from a single, theoretically elegant base or core, (iii) *sufficient efficiency*, aiming to execute fast enough for real-time applications and (iv) *generic cognition*, aiming to integrate both natural and artificial intelligence. These desiderata motivate Sigma's blending of graphical models (Koller & Friedman, 2009) with cognitive architectures, yielding a broadly capable and theoretically elegant base – referred to as the *graphical layer* – which supports a broad *cognitive layer* on top. The cognitive layer provides a language for adding skills and knowledge (Rosenbloom, 2009) on top of the architecture. Together, the cognitive and graphical layers have been shown to support a wide variety of capabilities – perception and decision making (Chen, et al., 2011), reinforcement learning (Rosenbloom, 2012), episodic memory & learning (Rosenbloom, 2014) etc. – as demanded by the goal of grand unification.

Thesis: *Basing the Sigma cognitive architecture on factor graphs enables the fusion of speech, language and cognition – a combination of symbolic and sub-symbolic capabilities. Speech and language can be added on top of the architecture as knowledge with the architecture enabling the fusion that brings various sources of knowledge into play during speech and language processing, with each capability aiding the other*.

This thesis proposes the integration of incremental speech processing with language understanding and cognition. Speech signal obtained from a typical speech front end shall be combined with linguistic knowledge in the form of phonetic and semantic knowledge sources with cognition selecting the most likely word incrementally. This integration is deemed *supraarchitectural* – i.e. on top of the architecture – because the capability is specified as knowledge on top of the general set of mechanisms that constitute the Sigma architecture.

In addition to integrating various supraarchitectural capabilities, using multiple interacting supraarchitectural capabilities together to yield more complex behaviors is also an important requirement for cognitive architectures (Rosenbloom, 2015). Such interoperation is not always easy or successful, as evidenced by the difficulties encountered in integrating declarative learning in Soar (Rosenbloom, Newell, & Laird, 1991) with other capabilities (Rosenbloom, 2015). Thus, an important evaluation of the usefulness of a supraarchitectural capability, in a cognitive architecture setting, is to study the feasibility of using it in conjunction with another capability to yield more complex behavior. An extension and partial evaluation of the integration of speech processing is proposed by reusing the speech capability, in conjunction with a language understanding capability to construct a suitably chosen discourse capability. Linguistic input from the speech capability shall be converted into meaning using a Natural Language Understanding (NLU) scheme such as the one proposed in (DeVault, Sagae, & Traum, 2011). Cognition shall select the best meaning, in conjunction with the state and goals of a discourse task, plus possibly additional knowledge, to choose the next response. Additionally, because language understanding and discourse management are supraarchitectural capabilities, they can influence the speech capability using state information, derived potentially from goals of the task, input from other modalities, emotional appraisals etc. This continuous bidirectional flow of information between speech, language understanding result in a tight coupling of incremental speech processing, language understanding and cognition gives rise to a form of adaptive language model that is both *dynamic* – able to change every 50 msec – and *open* – to a range of knowledge sources, linguistic and otherwise. This type of dynamic language model can be assessed in the context of a discourse task, where incremental speech processing that can exploit context or potentially react to other modalities in a continuous and timely manner is a key requirement.

Sigma has demonstrated the ability to process a form of speech signal to perform isolated word recognition, in an incremental fashion (Joshi, Rosenbloom, & Ustun, 2014). Additionally, much of the structure needed for speech recognition was constructed automatically by Sigma, using templates added in service of SLAM and RL. The parameters required for this task were learnt using Sigma's gradient descent learning. Subsequently, a more natural form of speech processing was implemented using the TIMIT database (Zue, Seneff, & Glass, 1990) to perform continuous speaker-independent phone recognition. Sigma has previously shown the ability to learn distributed semantic vectors (Ustun, Rosenbloom, Sagae, & Demski, 2014).

References

Chen, J., Demski, A., Han, T., Morency, L., Pynadath, P., Rafidi, N., & Rosenbloom, P. S. (2011). Fusing symbolic and decision-theoretic problem solving + perception in a graphical cognitive architecture. *Proceedings of the Second International Conference on Biologically Inspired Cognitive Architectures*.

DeVault, D., Sagae, K., & Traum, D. (2011). Incremental interpretation and prediction of utterance meaning for interactive dialogue. *Dialogue & Discourse, 2*((1)), 143-170.

Joshi, H., Rosenbloom, P. S., & Ustun, V. (2014). Isolated word recognition in the Sigma cognitive architecture. *Biologically Inspired Cognitive Architectures*(10), 1-9.

Koller, D., & Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques.* Cambridge, MA: MIT Press.

Laird. (2012). *The Soar Cognitive Architecture.* Cambridge, MA: MIT Press.

Langley, P., Laird, J. E., & Rogers, S. (2009). Langley, Pat, John E. Laird, and Seth Rogers. *Cognitive Systems Research, 10*(2), 141-160.

Lebiere, C., O'Reilly, R. C., Jilk, D. J., Taatgen, N., & Anderson, J. R. (2008). The SAL Integrated Cognitive Architecture. *AAAI Fall Symposium: Biologically Inspired Cognitive Architectures*, 98-104.

O'Reilly, R. C., Hazy, T. E., & Herd, S. A. (2012). The leabra cognitive architecture: how to play 20 principles with nature and win. *The Oxford Handbook of Cognitive Science*.

O'Reilly, R. C. (1998). Six principles for biologically based computational models of cortical cognition. *Trends in cognitive sciences, 2*(11), 455-462.

Rosenbloom. (2009). A graphical rethinking of the cognitive inner loop. *Proceedings of the IJCAI International Workshop on Graphical Structures for Knowledge Representation and Reasoning*.

Rosenbloom. (2012). Deconstructing reinforcement learning in Sigma. *Proceedings of the 5th Conference on Artificial General Intelligence*.

Rosenbloom. (2014). Deconstructing episodic memory and learning in Sigma. *Proceedings of the 36th Annual Conference of the Cognitive Science Society*.

Rosenbloom. (2015). Supraarchitectural capability integration: From Soar to Sigma. *Proceedings of the 13th International Conference on Cognitive Modeling.*

Rosenbloom, P. S., Newell, A., & Laird, J. E. (1991). Towards the knowledge level in Soar: The role of the architecture in the use of knowledge. *Architectures for Intelligence: The Twenty-second Carnegie Mellon Symposium on Cognition*, 75-111.

Sun, R. (2006). The CLARION cognitive architecture: Extending cognitive modeling to social simulation. In R. e. Sun, *Cognition and Multi-Agent Interaction* (pp. 79-99). New York: Cambridge University Press.

Sun, R., Merrill, E., & Peterson, T. (2001). *From implicit skills to explicit knowledge: A bottom-up model of skill learning.* (Vol. 25). Cognitive Science.

Ustun, V., Rosenbloom, P. S., Sagae, K., & Demski, A. (2014). Distributed vector representations of words in the Sigma cognitive architecture. *Artificial General Intelligence*, 196-207.

Zue, V., Seneff, S., & Glass, J. 9.-3. (1990). Speech database development at MIT: TIMIT and beyond. *Speech Communication, 9*(4), 351-356.