Cognitive Memory Modeling for Interactive Systems in Dynamic Environments

Felix Putze and Tanja Schultz

Karlsruhe Institute of Technology (KIT), Germany,
{felix, tanja}@ira.uka.de,
WWW home page: http://csl.ira.uka.de

Abstract. This paper presents our ongoing work on cognitive modeling components designed for applications in adaptive dialog systems for dynamic environments. We believe that insight into which information the user currently processes is necessary to build dialog systems that deal with dynamic environments like a car in which the user is exposed to multiple distractions and stimuli. We present our approach to model the user's imperfect memory during interaction, inspired by psychologically sound cognitive architectures.

1 Motivation & Applications

Spoken dialog systems (SDS) have matured to a point where they find their way into many real-world applications. However, their application in very dynamic scenarios remains an open and very challenging task. For example, while SDS offer an eyes-free and hands-free control for in-car information applications, the parallel driving task uses the user's cognitive capacity so we can no longer assume to deal with a fully attentive and perfect interaction partner as in more static environments. Therefore, we need to integrate components in our dialog systems that are able to explicitly model, predict, and cope with this imperfect user to ensure a seamless and successful dialog experience. In this work, we concentrate on the user's memory from which his beliefs and goals are derived. Unfortunately, most current user models and user simulations implicitly assume the user's memory to be equivalent to the (omniscient) discourse model (see [Schatzmann et al., 2006] for an overview). An explicit *memory model* aims for a more detailed representation of human memory by modeling an individual strength of activation for each chunk of information.

Cognitive modeling components are useful for a large variety of applications in the context of interaction management. The task of **System Utterance Selection** is to give appropriate information to the user, dependent on the context. In systems with a high degree of system initiative or task orientation, the selection of the right information is straight-forward, given enough specification from the user. However, in more free interaction scenarios, e.g. for a conversational in-car tourguide system, the choice becomes more ambigious. A memory model helps to identify the most relevant information bits and may also work with less direct user intervention due to memory inference.

In [Putze and Schultz, 2009], we propose to employ the *urge* mechanism from the PSI cognitive architecture to model the strength of different desires and needs of the user. This mechanism is useful to weight different (potentially adversarial) goals and associated system actions against each other. One important aspect is the **Competence Urge**, i.e. the desire to acquire additional domain information. We estimate this urge by inspecting active items in the memory model.

User Simulation is relevant as an evaluation technique as well as for dialog strategies trained with Reinforcement Learning. Current statistical user simulation techniques [Schatzmann et al., 2006] lack coherent and motivated behavior that is necessary to create realistic interaction sessions, especially when the dialog becomes less task-driven. A memory model integrated in the user model helps to improve on this: On the basis of the most active items in memory, the simulated user selects its actions to signal its preferences to the system.

Many other applications are thinkable. Memory modeling can form a basis for modeling the **human speech understanding** process (to account for potential misunderstandings) or to **support the speech recognizer** with an increased weight on the items currently activated in the user's mind.

2 From Cognitive Architectures to User Memory Modeling in Dialog Systems

Cognitive architectures try to accommodate all important phenomena and interrelations of human cognition. Famous examples are ACT-R [Anderson et al., 2004] and PSI [Dörner, 2002]. We believe that these ambitious cognitive architectures are not yet mature enough for application in unconstrained real-world tasks, but they offer components which are already useful to model certain aspects of human cognition. We specialize them for our needs while maintaining the possibility of integrating more components later. One central component present in all cognitive architectures is the *memory model*. It represents the human's short- and long-term memory, i.e. declarative knowledge which can be activated and reinforced by internal and external stimuli. ACT-R for example represents knowledge in form of *chunks*, i.e. typed bits of information that are composed of so-called slots that are either filled with atomic information (e.g. a number) or other chunks. Instead of the original memory model of ACT-R, we use the improved LTM^c [Schultheis et al., 2006] version. Here, chunks are represented as nodes in a semantic network, connected by edges along which activation spreads between nodes. Associated with each chunk is an activation value which describes the degree of availability and predicted usefulness of the corresponding chunk in the future. Activation is determined by a formula (for details, see [Anderson et al., 2004]) which consists of three summands: The first part of the equation describes the *basic activation*, which increases with the number and recency of past presentations of the chunk. The second part, *spreading activation*, is activation that spreads from associated (e.g. similar or related) chunks. The third component is *noise*. At runtime, external components send presentations and requests of chunks to the model. Presentations change activation scores

directly (basic activation) and indirectly (spreading activation) while requests either retrieve the activation of a specific chunk or the set of the most activated chunks in memory.

When using a memory model as a tool to identify the most relevant bits of information (e.g. to select the most appropriate system utterance), a first approach would probably use the activation value as an indication of relevance. From the definition above, we see that the activation value of chunks actually summarizes multiple influences and time scales of memory effects: The basic activation can be interpreted as the result learning from frequent presentations of the corresponding chunk. On the other hand, the volatile spreading activation results not from direct presentation but from associations to other activated nodes. When selecting the next system action, we may not want to present information which is highly activated due to frequent and recent presentation causing a high basic activation. On the other hand, chunks which are not activated at all will be uninteresting or unexpected for the user. We therefore define the concept of *importance*, which measures the potential benefit of presenting the respective item to the user. Importance in its simplest form denotes the ratio between spreading activation and basic activation. This marks a chunk as especially important for presentation if it was activated through associated nodes but not directly presented.

We now describe the development of the memory model in our case of an in-car tourguide system. We begin with a network with no activation besides noise. While the user drives on a route in our driving simulator [Putze and Schultz, 2009], stimuli for an increase of activation come from the *environment model* which represents events outside the driver cabin that influence the user. Examples are events which trigger the activation of certain chunks. They are based on tracking the user's position on his route, fetching nearby points-of-interest (POI) from our database[1] and activating them in the memory model. They can also be triggered by direct user intervention, asking for information on the POI. This activation spreads to connected nodes for which the spreading activation (but not the basic activation) rises and therefore their importance. To select from the set of system utterances, we tag them with the chunks that are activated by it (e.g. 'This church [church] was built in [built_in] the 18th century [18th_century]'). Then, we iterate over all utterances and select the one which maximizes the accumulated importance of the corresponding memory chunks. For example, the system will prefer the utterance from the previous example over the utterance 'This church [church] is dedicated [dedicated] to St. Mary [St_Mary]', when the earlier presentation of the [18th_century]-chunk spreaded to the [built_in] node.

---

[1] The database currently contains five POIs for a drive of ten minutes. Each POI corresponds to 10-20 chunks. There are alternative utterances (e.g. varying in complexity) covering the same chunks.
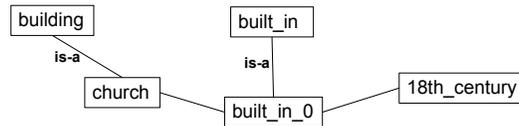
Fig. 1. Excerpt of a memory network in LTM^c. Activation of `church` by the environment model will spread to neighboring nodes.

3 Future Work

The next step in the development process is systematic evaluation. We rely on established cognitive modeling components which are psychologically sound and already tested in many applications. To go beyond that, cognitive components for interactive systems have to be evaluated directly in interactive situations and in concert with all other components. Currently ongoing are simulation experiments where in a Reinforcement Learning framework both the employed system and the user simulation are based on the concepts presented here. Successful generation of cognitive strategies will be a first proof for their usefulness and tentative results are very promising. The next step, real user tests, will determine how well those strategies and therefore the involved cognitive modeling components can be transfered to the real world.

Another step is of course the integration and interconnection of more cognitive components. Currently, the memory model described here as well as a motivational urge model [Putze and Schultz, 2009] already exist. Both could be combined with other components like the procedural module from ACT-R (to model activation between external stimuli) or the emotion modeling from PSI (to model the effect of emotions on memory).

References

[Anderson et al., 2004] Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111:1036–1060.

[Dörner, 2002] Dörner, D. (2002). *Die Mechanik des Seelenwagens*. Huber, Bern.

[Putze and Schultz, 2009] Putze, F. and Schultz, T. (2009). Towards cognitive dialog systems. In *1. Berliner Fachtagung Biophysiologische Interfaces*.

[Schatzmann et al., 2006] Schatzmann, J., Weilhammer, K., Stuttle, M., and Young, S. (2006). A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies. *The Knowledge Engineering Review*, 21.

[Schultheis et al., 2006] Schultheis, H., Barkowsky, T., and Bertel, S. (2006). Ltm-c - an improved long-term memory for cognitive architectures. In *Proceedings of the 7th International Conference on Cognitive Modeling*.