Tensor Memories

Volker Tresp, Yunpu Ma, Stephan Baier (fn.ln@siemens.com)
Siemens Corporate Technology, Otto-Hahn-Ring 6
Muenchen, Germany

Abstract

The hippocampal memory indexing theory of Teyler and DiScenna is one of the leading theories of memory formation. The main idea is that high-dimensional latent representations are formed in the hierarchical sensory processing layers of the brain as response to sensory inputs, and that these representations are linked to indices in the hippocampal area, forming a basis for an episodic memory. We extend the hippocampal memory indexing theory by including a decoder that extracts explicit information in the form of semantic triples using latent representations of entities and predicates. We demonstrate that, if a tensor model is used for explicit decoding, a semantic memory can be derived from episodic memory by a marginalization operation. Thus, our model supports the assumption that semantic memory is derived from episodic memory and that both rely on the same latent representations of generalized entities.

Keywords: Episodic memory; semantic memory; representation learning; perceptual decoding; tensor models

Introduction

The hippocampal memory indexing theory (HMIT) of Teyler and DiScenna is one of the leading theories of memory formations (Teyler & DiScenna, 1986; Teyler & Rudy, 2007). The main idea is that high-dimensional latent representations are formed in the hierarchical sensory processing layers of the brain as response to sensory inputs, and that these representations are linked to indices in the hippocampal area, forming a basis for an episodic memory. One of the interesting features of the theory is that new episodic memories can be stored without affecting already stored memories, thus avoiding memory interference. The theory is subsymbolic: it shows how past memories can be recovered by re-activating the time index which then restores the same sensory processing patterns that were originally generated by the sensory inputs. However, the theory does not address a number of essential aspects of episodic memory: first, episodic memory is an explicit memory, which means that individuals can *describe* past episodic memories. A second issue is the relationship between episodic memory and semantic memory. How can semantic memory ("things we know") be derived from episodic memory ("things we remember")? In this paper, we present a mathematical model, based on tensor decompositions, which implements explicit decoding of sensory inputs and, at the same time, is instrumental in forming explicit episodic and semantic memories. Our tensor memories are closely related to statistical models of knowledge graphs, which have found a number of technical applications (Nickel, Murphy, Tresp, & Gabrilovich, 2015).

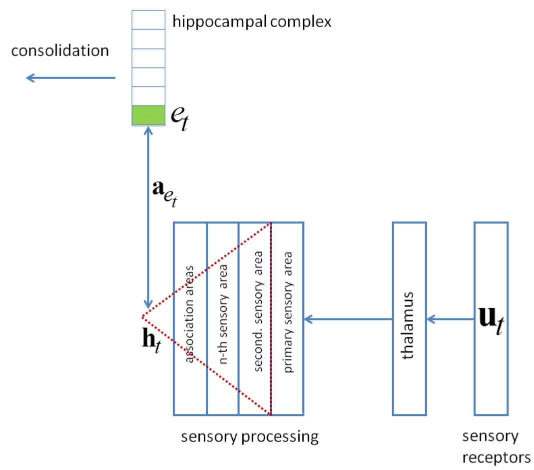

Figure 1: Main components of the HMIT.

The Hippocampal Memory Indexing Theory

Figure 1 summarizes some of the main components of the HMIT. It is shown how information from sensory input is relayed through the thalamus (except olfaction), a part of the forebrain, and then forms the input to the hierarchical sensory processing layers. A sensory response at time t, denoted as $\mathbf{h}_t$, typically activates most areas in the early processing steps, and is more specific at the later processing steps, relying on few higher order features (this property is represented by the red triangle). If an episode is worth remembering, e.g., because it is significant, novel, or emotional, an index e_t is formed in the hippocampus and $\mathbf{h}_t$ is stored as a connection pattern $\mathbf{a}_{e_t} = \mathbf{h}_t$. If a past episode is recalled, the corresponding index e_t is re-activated (for example by a partial sensory pattern or by other means), and then $\mathbf{a}_{e_t}$ is approximately restored in the sensory processing layers, thus providing a sensory impression of the past episode. The HMIT provides a detailed hypothesis on how the index is generated in the hippocampal area. In a memory consolidation process, the index e_t might also find a representation in the neocortex. Current thinking is that consolidation happens during non-REM sleep, possibly by a form of replay. The representation in the cortex might be index-based and, in addition or instead, might have complementary memory representations, as well (Kumaran, Hassabis, & McClelland, 2016).

Explicit Episodic Memories

The episodic memory in the HMIT does not really provide a means of extracting declarative information: a recall simply re-activates the hierarchical sensory layers. In this paper we assume that an explicit decoding produces "triples in time" as (s, p, o, t) where s is the subject (head entity), p is the predicate, o is the object (tail entity), and t is the time step. $(e_t, \mathbf{a}_{e_t})$, but also that each entity i is represented by the pair $(e_i, \mathbf{a}_{e_i})$, and a predicate k is represented by the pair $(e_k, \mathbf{a}_{e_k})$. It is typically assumed that indices for entities e_i are also formed in the hippocampal region and are later consolidated in neocortex, where they form topological maps (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016). For the decoding step, we assume a joint probabilistic model based on a tensor decomposition. An example is a Tucker model with $P(s, p, o, t) \propto f^{epi}(s, p, o, t)$ and where

$$f^{epi}(s, p, o, t) =$$

$$\sum_{r_1=1}^{\tilde{r}_1} \sum_{r_2=1}^{\tilde{r}_2} \sum_{r_3=1}^{\tilde{r}_3} \sum_{r_4=1}^{\tilde{r}_4} a_{e_s,r_1}\, a_{e_p,r_2}\, a_{e_o,r_3}\, a_{e_t,r_4}\, g(r_1, r_2, r_3, r_4). \quad (1)$$

Here, $g(r_1, r_2, r_3, r_4) \in \mathbb{R}$ are elements of the core tensor $\mathcal{G} \in R^{\tilde{r}_1 \times \tilde{r}_2 \times \tilde{r}_3 \times \tilde{r}_4}$, where $\tilde{r}_1, \tilde{r}_2, \tilde{r}_3, \tilde{r}_4$ are the ranks. All parameters are constrained to be nonnegative.

Declarative decoding generates a set of (s, p, o) triples from $P(s, p, o|t)$ where $\mathbf{a}_{e_t}$ is either generated from sensory input ($\mathbf{a}_{e_t} = \mathbf{h}_t$) or from re-activating the corresponding past time index e_t. For example, a visual scene is decoded as a triple set which describes that scene, as in (Lu, Krishna, Bernstein, & Fei-Fei, 2016). We can also calculate marginal and other conditional probabilities in biologically plausible ways (Tresp et al., 2015).

Semantic Memory

It is generally assumed that episodic memory is the gateway to semantic memory, but the precise relationship is still unclear. In our model we can generate semantic memory by simply marginalizing episodic memory. Thus if *(Jack, diagnosed-With, Diabetes, January 10, 2016-01-10)* is part of episodic memory, the derived semantic memory is *(Jack, diagnosed-With, Diabetes)*. Since nonnegative tensor models are sum-product nets, marginalization is trivial. One can also derive $P(o|s, p)$ answering queries, e.g., as "what *(=o)* does Jack *(=s)* like *(=p)*?". The model can also accommodate for start-date and end-date of a fact, such as a disease which was first diagnosed and then cured, by introducing a representation for negation, i.e., a representation for the end-date. Details on the mathematical operations in the model can be found in (Tresp et al., 2015).

First experimental results for a link prediction task on the ICEWS knowledge base are shown in Figure 2. We show recall scores for the training and the test set as a function of the rank of the model.

We report results for the semantic memory ("Semantic"), for the semantic memory projected from the episodic memory with considering only representation of starting point ("Start"), and for the projected semantic memory with considering representations of both stating and terminal points ("Start-End"). The figures confirm that we can obtain a semantic memory by projecting the episodic memory in a proper way, supporting the hypothesis that episodic memory is a "gateway" to semantic memory.

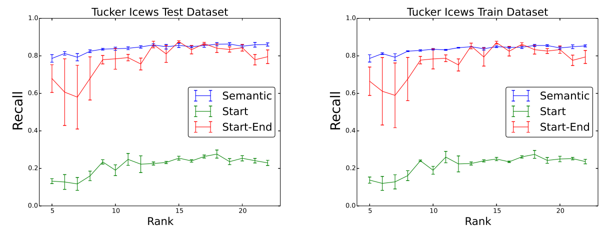

Figure 2: Rank-dependent recall scores for training and testing data for models with nonnegative parameters.

Discussion

We have discussed how semantic and episodic adjacency tensors and their decompositions might serve as models for human explicit memories. Our models only use biologically plausible constructs, such as latent representations of entities and time steps. Our technical memories are approximate and provide a basis for inductive inference. In our experiments we have demonstrated recall of memories but also the generalization to new facts. Of course, generalization also brings the danger of over generalization, or false memories. Note that, although an index might only involve few neurons (Quiroga, 2012), the index is never activated alone, but is activated by, and activates, the associated latent representations in the sensory processing layers.

References

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*.

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What learning systems do intelligent agents need? *Trends in Cognitive Sciences*, *20*(7), 512–534.

Lu, C., Krishna, R., Bernstein, M., & Fei-Fei, L. (2016). Visual relationship detection with language priors. In *ECCV*.

Nickel, M., Murphy, K., Tresp, V., & Gabrilovich, E. (2015). A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*.

Quiroga, R. Q. (2012). Concept cells: the building blocks of declarative memory functions. *Nat Rev Neurosci*, *13*(8).

Teyler, T. J., & DiScenna, P. (1986). The hippocampal memory indexing theory. *Behavioral neuroscience*.

Teyler, T. J., & Rudy, J. W. (2007). The hippocampal indexing theory and episodic memory. *Hippocampus*.

Tresp, V., Esteban, C., Yang, Y., Baier, S., Ma, Y., & Krompaß, D. (2015). Learning with memory embeddings. *arXiv preprint arXiv:1511.07972*.