Learning with Memory Embeddings

Volker Tresp, Cristóbal Esteban, Yinchong Yang,
Stephan Baier and Denis Krompaß
Siemens AG and Ludwig Maximilian University of Munich
81739 Munich, Germany

Abstract

Embedding learning, a.k.a. representation learning, has been shown to be able to model large-scale semantic knowledge graphs. A key concept is a mapping of the knowledge graph to a tensor representation whose entries are predicted by models using latent representations of generalized entities. In recent publications the embedding models were extended to also consider temporal evolutions, temporal patterns and subsymbolic representations. In this paper we map embedding models, which were developed purely as solutions to technical problems for modelling temporal knowledge graphs, to various cognitive memory functions, in particular to semantic and concept memory, episodic memory and sensory memory. A hypothesis that arises out off this work is that mutual information exchange can be achieved by sharing or coupling of distributed latent representations of entities across different memory functions, in accordance, e.g., with cognitive theories about the mutual support of episodic and semantic memory.

1 Introduction

Embedding learning, a.k.a. representation learning, has been shown to be able to model large-scale semantic knowledge graphs [49, 12, 20, 51]. A key concept is a mapping of the knowledge graph to a tensor representation whose entries are predicted by models using latent representations of generalized entities. In recent publications the embedding models were extended to also consider temporal evolutions, temporal patterns and subsymbolic representations [24, 25]. These extended models were used successfully to predict clinical events like procedures, lab measurements, and diagnoses. In this paper, we attempt to map these embedding models, which were developed purely as solutions to technical problems, to various cognitive memory functions, in particular to semantic and concept memory, episodic memory and sensory memory. We also make an analogy between a predictive model, which uses entity representations derived in memory models, to working memory. Our approach follows the tradition of Latent Semantic Analysis, which is a classical representation learning approach that on the one hand has found a number of technical applications and on the other hand could be related to cognitive semantic memories [43, 42].

Cognitive memory functions are typically classified as *long-term* or *short-term* memory, where long-term memory has the subcategories *declarative* memory and *non-declarative* memory and the short term memory has the subcategories *sensory* memory and *working* memory [22, 4, 9, 59, 19, 27]. There is evidence that these main cognitive categories are partially dissociated from one another in the brain, as expressed in their differential sensitivity to brain damage [27]. However, there is also evidence indicating that the different memory functions are not mutually independent and support each other [36, 29]. A hypothesis that arises out off this work is that mutual information exchange can be achieved by sharing or coupling of distributed latent representations of entities across different memory functions, in accordance, e.g., with cognitive theories about the mutual support of episodic and semantic memory [59, 6, 64, 67].

2 Technical Models for Declarative Long-Term Memory

Declarative memories are long-term memories that work on a semantic abstraction level. Technical and human declarative memories can be divided into *concept memory* which stores information about the concepts in the world and their hierarchical organization, *semantic memory* which stores general world knowledge about entities, and *episodic memory* which stores events of prior personal experiences [62, 63, 64, 27]

2.1 A Semantic Knowledge Graph Model

A technical realization of a semantic memory is a knowledge graph (KG) which is a triple-oriented knowledge representations. Popular large-scale KGs are DBpedia [5], Yago [60], Freebase [11], and the Google Knowledge Graph [57].

Here we consider a slight extension to the subject-predicate-object triple form by adding the value in the form (e_s, e_p, e_o; *Value*) where *Value* is a function of s, p, o and, e.g., can be a Boolean variable (*True* or *1*, *False* or *0*) or a real number. Thus *(Jack, likes, Mary; True)* states that Jack likes Mary. Note that e_s and e_o represent the entities for subject index s and object index o. To simplify notation we also consider e_p to be a generalized entity associated with predicate type with index p. We encode attributes also as triples, mostly to simplify the discussion.

We now consider an efficient representation of a KG. With this representation, it is also possible to generalize from known facts to new facts (inductive inference). First, we introduce the three-way semantic adjacency tensor $\underline{X}$ where the tensor element $x_{s,p,o}$ is the associated *Value* of the triple (e_s, e_p, e_o). Here $s = 1, \ldots, S$, $p = 1, \ldots, P$, and $o = 1, \ldots, O$. One can also define a companion tensor $\underline{\Theta}$ with the same dimensions as $\underline{X}$. It contains the natural parameters of the model and the connection to $\underline{X}$ for Boolean variables is $P(x_{s,p,o}|\theta_{s,p,o}) \sim \mathrm{sig}(\theta_{s,p,o})$, where $\mathrm{sig}(arg) = 1/(1 + \exp(-arg))$ is the logistic function (Bernoulli likelihood) . If $x_{s,p,o}$ is a real number then we can use a Gaussian distribution with $P(x_{s,p,o}|\theta_{s,p,o}) \sim \mathcal{N}(\theta_{s,p,o}, \sigma^2)$.

The key concept in embedding learning is that each entity e has an r-dimensional latent vector representation $\mathbf{a} \in \mathbb{R}^r$. In particular, the embedding approaches used for modeling KGs assume that[1]

$$\theta^{semantic}_{s,p,o} = f^{semantic}(\mathbf{a}_{e_s}, \mathbf{a}_{e_p}, \mathbf{a}_{e_o}).$$

The functions $f^{semantic}$ can be modeled. e.g., via feedforward neural networks or as a tensor factorization [51]. In case of a KG with a Bernoulli likelihood, $\mathrm{sig}(\theta^{semantic}_{s,p,o})$ represents the confidence that the *Value* of the triple (e_s, e_p, e_o) is true.

Latent representation approaches have been used very successfully to model large KGs, such as the Yago KG, the DBpedia KG and parts of the Google KG. It has been shown experimentally that models using latent factors perform well in these high-dimensional and highly sparse domains. Since an entity has a unique representation, independent of its role as a subject or an object, the models permits the propagation of information across the KG. For example if a writer was born in Munich, the model can infer that the person is born in Germany and probably writes in the German language [49, 50]. Stochastic gradient descent (SGD) is typically being used as an iterative approach for finding both optimal latent representations and optimal parameters in $f^{semantic}(\cdot)$ [51, 41]. For a recent review, please consult [51].

To answer the query, "Who is married to *Jack*", we can execute $e_o = \arg\max_e \theta^{semantic}_{Jack,marriedTo,e}$. Due to the approximation, $\mathrm{sig}(\theta^{semantic}_{Jack,marriedTo,e})$ might be smaller than one for the true spouse. The approximation also permits inductive inference: We might get a large $\mathrm{sig}(\theta^{semantic}_{Jack,marriedTo,e})$ also for entities e that are *likely* to be married to *Jack* and $\mathrm{sig}(\theta^{semantic}_{s,p,o})$ can, in general, be interpreted as a confidence value for the triple (s, p, o). More complex queries on semantic-models involving existential quantifier are discussed in [40].

A concept memory would technically correspond to classes with a hierarchical subclass structure. In [48, 47] such a structure was learned from the latent representation via hierarchical clustering. In KGs, a hierarchical structure is described by *type* and *subclass* relations.

[1] An alternative representation is $\theta_{s,p,o} = f^{semantic}_p(\mathbf{a}_{e_s}, \mathbf{a}_{e_o})$ with a separate function for each predicate type but to simplify the discussion we will not discuss this version further in the paper.

Latent representations for modeling semantic memory functions have a long history in cognitive modeling, e.g., in latent semantic analysis [42] which is restricted to attribute-based representations. Latent clustering and topic models [37, 66, 1] are extensions toward multi-relational domains and use discrete latent representations. See also [44, 30, 31]. Spreading activation is the basis of the Teachable Language Comprehender (TLC), which is a network model of semantic memory [15]. Associate models are the symbolic ACT-R [2, 3] and SAM [54]. [52] explores holographic embeddings with representation learning to model associative memories. Connectionists memory models are described in [35, 45, 14, 38, 32, 33].

2.2 An Event Model for Episodic Memory

Whereas a KG model reflects the state of the world, e.g, of a clinic and its patients, observations and actions describe factual knowledge about discrete events, which, in our approach, are represented by an event tensor. In a clinical setting, events might be a prescription of a medication to lower the cholesterol level, the decision to measure the cholesterol level and the measurement result of the cholesterol level; thus events can be, e.g., actions, decisions and measurements.

The episodic event tensor is a four-way tensor $\underline{Z}$ where the tensor element $z_{s,p,o,t}$ is the associated *Value* of the quadruple (e_s, e_p, e_o, e_t). We model

$$\theta^{event}_{s,p,o,t} = f^{event}(\mathbf{b}_{e_s}, \mathbf{b}_{e_p}, \mathbf{b}_{e_o}, \mathbf{b}_{e_t}) \quad \text{or} \quad \theta^{event}_{s,p,o,t} = f^{event}(\mathbf{b}_{e_p}, \mathbf{b}_{e_o}, \mathbf{b}_{e_{s,t}})$$

where $\mathbf{b}_{e_s}, \mathbf{b}_{e_p}, \mathbf{b}_{e_o}, \mathbf{b}_{e_t}$ and $\mathbf{b}_{e_{s,t}}$ are latent representations for subject, predicate, object, time and subjectTime. Note that we have added a representation for the time of an event by introducing the generalized entity e_t with latent representation $\mathbf{b}_{e_t}$. This latent representation compresses all events that happen at time t. In addition we have introduced the generalized entity $e_{s,t}$ with the latent representation $\mathbf{b}_{e_{s,t}}$ which represents all events of subject entity e_s at time t.

As examples, the agent can recall "Who did I meet last week?" by $e_o = \arg\max_e \theta^{event}_{Myself,meet,e,LastWeek}$ and "When did I meet Jack?" by $e_t = \arg\max_e \theta^{event}_{Myself,meet,Jack,e}$.

Examples from our clinical setting would be: *(Jack, orderBloodTest, Cholesterol, Week34; True)* for the fact that a cholesterol blood test was ordered in week 34 and *(Jack, hasBloodTest, Cholesterol, Week34; 160)* for the result of the blood test. In many applications the most important latent representation is $\mathbf{b}_{e_{s,t}}$ which can be thought of as a compressed representation of all events that happened to the subject (e.g., the patient) at time t. Note that we consider episodic memory over different subjects, predicates and objects; thus semantic memory can represent an extensive event context!

An event model can be related to the cognitive concept of an *episodic memory*. Episodic memory represents our memory of experiences and specific events in time in a serial form (a "mental time travel"), from which we can reconstruct the actual events that took place at any given point in our lives [58][2]. In contrast to semantic memory, it requires recollection of a prior experience [63]. A part of the episodic memory is the *autobiographical memory* which stores autobiographical events of an agent on a semantic abstraction level [27]. One might argue that in a true autobiographic model, the subject is in general the agent.

In our technical model, some events might be copied to the KG model [24, 25]. For example the event model might record a diagnosis which then becomes a fact in the KG. A similar affect is part of some cognitive models. It is sometimes assumed that the semantic memory is derived from the episodic memory, in that we learn new facts or concepts from our experiences, and the episodic memory is considered to support and support semantic memory. A gradual transition from episodic to semantic memory can take place, in which episodic memory reduces its sensitivity and association to particular events, so that the information can be generalized as semantic memory. Thus some theories speculate that episodic memory may be the "gateway" to semantic memory [59, 6, 67].[3]

[2] http://www.human-memory.net/types_episodic.html

[3] Declarative memories are encoded by the hippocampus and other areas the medial temporal lobe (MTL) of the brain (in particular, generation and storage of episodic memory); they are consolidated and stored in the temporal cortex (anterolateral temporal lobe) and might be distributed across a large number of brain areas (in particular semantic memory) [16, 56, 29, 10].

3 Sensor Inputs

3.1 A Sensory Memory

We realize a sensor memory by assuming a buffer that holds the last T realizations of some abstract sensory inputs at time t. The sensory tensor is a three-way tensor $\underline{U}_t$ where the tensor element $u_{q,\tau,t}$ is the associated *Value* of the triple (e_q, e_τ, e_t). e_q is a generalized entity for the q-th sensory channel, e_τ ($\tau = 0, \ldots, T$) specifies the temporal location in the buffer and e_t is a generalized entity representing the complete buffer at time t. Since the subject is always the agent we do not need to introduce a subject representation.

We get

$$\theta^{\textit{sensory}}_{q,\tau,t} = f^{\textit{sensory}}(\mathbf{c}_{e_q}, \mathbf{c}_{e_\tau}, \mathbf{c}_{e_t})$$

where $\mathbf{c}_{e_q}$, $\mathbf{c}_{e_\tau}$ and $\mathbf{c}_{e_t}$ are latent representations for the sensor channel e_q, temporal buffer position e_τ and time e_t. Latent components corresponds to complex channel-temporal patterns (chunks) whose amplitudes are determined by the components of $\mathbf{c}_{e_t}$.

In a technical application [25], the sensors measure, e.g., wind speed, temperature, and humidity at the location of a wind turbine and the sensory memory retains the measurements from t to $t - T$.[4]

In human cognition, sensory memory (milliseconds to a second) represents the ability to retain impressions of sensory information after the original stimuli have ended [62, 17, 27]. The transfer of sensory memory to short term memory is the first step in some memory models [4]. Sensory memory can be the basis for sequence learning and the detection of complex temporal patterns.

One can envision several short-term memory (STM) buffers for sensory input with different lengths T, and also an STM buffer for episodic memory, for which there is some cognitive evidence [36, 8].

3.2 Object Representation

Consider that the sensory memory (e.g., derived from visual or auditorial inputs) represents an entity e. The sensor channels might have been calculated by complex transformations based on some raw sensor input. For example, in some work on face recognition, the representation is calculated from sensor inputs by a complex normalization step followed by several neural network layers [61] performing complex perceptual (possibly multimodal) processing. In this case we can interpret e_t as the sensor impression of some object e_o and we can perform object recognition with $e_o = \arg\min_e \text{dist}(\mathbf{a}_e, \mathbf{c}_{e_t})$, where dist$(\cdot, \cdot)$ is an appropriate distance function.

4 Learning with Memory Embeddings and a Relation to Working Memory

The memory embeddings can be the basis for solving classification and prediction tasks. For example, in a clinical setting, it is important to know what should be done next (e.g., prediction of a medical procedure) or what event will happen next (e.g., prediction of a medical disagnosis).

For the predictive task we introduce a Markov model with a representation of the form

$$\theta^{\textit{Markov}}_{s,p,o,t} = f^{\textit{Markov}}(\textit{args})$$

where *args* is from the sets of latent representations from the semantic memory, the episodic memory or the sensory memory. In the medical application we are considering, we want to predict a diagnosis, e.g., *(Jack, hasDiagnosis, Diabetes, Today)* or a procedure *(Jack, orderLab, Bloodcount, Today)*. Arguments then are $\mathbf{a}_{e_s}, \mathbf{a}_{e_p}, \mathbf{a}_{e_o}$, which are the latent semantic representations of the patient, predicate and object from the semantic memory. In addition, we use $\mathbf{b}_{e_{s,t}}, \mathbf{b}_{e_{s,t-1}}, \ldots, \mathbf{b}_{e_{s,t-T}}$ from the event memory, which represents all events associated with the patient at times $t, ..., t - T$.

Prediction of events and actions on a semantic level can be considered one of the important functions of a cognitive working memory [53]. Working memory is the limited-capacity store for retaining information over the short term and performing mental operations on the contents of this store. As in our predictive Markov model, the contents of working memory could either originate from sensory

[4] In this application we also use a representation for a subject, since we consider a network of wind turbines.

input, episodic memory, or from semantic memory [27]. Cognitive models of working memory are described in [7, 8, 18, 23] and computational models are described in [21, 26, 13, 36, 53, 34, 65, 28].[5]

5 Conclusions and Discussion

We have discussed how a number of technical memory functions can be realized by representation learning and we have made the connection to cognitive memory functions. More details on the cost functions and the learning procedures can be found in [24, 25] where we also present successful applications to clinical decision modeling, sensor network modeling and recommendation engines. The predictive Markov model shares representations from the semantic, episodic and sensory memory models. Thus one might speculate that an entity should have a unique latent representation across different memory functions or that representations should at least not be independent and should be coupled. E.g., we might require that $\mathbf{b}_{e_s} = \mathbf{a}_{e_s}$ or $\mathbf{b}_{e_s} \approx \mathbf{a}_{e_s}$.[6] Shared representations lead to a global propagation of information across all memory functions during stochastic gradient descent learning [49]. In addition to the latent representations, the models contain parameters (e.g., neural network weights) in the memory models and in the predictive Markov model. One can make a link between those parameters and an *implicit (non-declarative) long-term memory* [55].

The different memory functions are realized by feedforward systems, given the latent representations of the entities as inputs. In contrast, a query typically requires iterative solutions to find the best matches for entities. In [24, 25] we also discussed how the latent representations of new entities and additional time steps can efficiently be calculated.

The cognitive plausibility of the technical solutions can be increased by requiring non-negativity of factors, which typically also improves interpretability and increases sparseness [31, 46, 39].

Note that in contrast to the knowledge graph, where an entity is represented by a single node, in embedding learning an entity has a distributed representation in form of multiple latent components.

References

[1] Edoardo M. Airoldi, David M. Blei, Stephen E. Fienberg, and Eric P. Xing. Mixed Membership Stochastic Blockmodels. *Journal of Machine Learning Research*, 9:1981–2014, September 2008.

[2] John R Anderson. *The architecture of cognition*. Psychology Press, 1983.

[3] John R Anderson, Michael Matessa, and Christian Lebiere. ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, 12(4):439–462, 1997.

[4] Richard C Atkinson and Richard M Shiffrin. Human memory: A proposed system and its control processes. *The psychology of learning and motivation*, 2:89–195, 1968.

[5] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. DBpedia: A Nucleus for a Web of Open Data. In *The Semantic Web*, Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2007.

[6] Alan Baddeley. Cognitive psychology and human memory. *Trends in neurosciences*, 11(4):176–181, 1988.

[7] Alan Baddeley. Working memory. *Science*, 255(5044):556–559, 1992.

[8] Alan Baddeley. Working memory: theories, models, and controversies. *Annual review of psychology*, 63: 1–29, 2012.

[9] Frederic C Bartlett. *Remembering: A study in experimental and social psychology*, volume 14. Cambridge University Press, 1995.

[10] Jeffrey R Binder and Rutvik H Desai. The neurobiology of semantic memory. *Trends in cognitive sciences*, 15(11):527–536, 2011.

[11] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 1247–1250. ACM, 2008.

[5] Considering the localization of the function, there is an emerging consensus that most working memory tasks recruit a network of prefrontal cortex and parietal areas but also involve other areas (basal ganglia and amygdala) [27].

[6] An interdependence of different long and short term memory functions is supported by fMRI studies [36]. Similarly, semantic and episodic memories support one another [29].

[12] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating Embeddings for Modeling Multi-relational Data. In *Advances in Neural Information Processing Systems 26*. 2013.

[13] Neil Burgess and Graham Hitch. Computational models of working memory: putting long-term memory into context. *Trends in cognitive sciences*, 9(11):535–541, 2005.

[14] Gail A Carpenter. Neural network models for pattern recognition and associative memory. *Neural networks*, 2(4):243–257, 1989.

[15] Allan M Collins and Elizabeth F Loftus. A spreading-activation theory of semantic processing. *Psychological review*, 82(6):407, 1975.

[16] Allan M Collins and M Ross Quillian. Retrieval time from semantic memory1. *Cognitive psychology: Key readings*, page 395, 2004.

[17] Max Coltheart. Iconic memory and visible persistence. *Perception & psychophysics*, 27(3):183–228, 1980.

[18] Nelson Cowan. *Attention and memory*. Oxford University Press, 1997.

[19] Nelson Cowan. What are the differences between long-term, short-term, and working memory? *Progress in brain research*, 169:323–338, 2008.

[20] Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. Knowledge Vault: A Web-scale Approach to Probabilistic Knowledge Fusion. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2014.

[21] Daniel Durstewitz, Jeremy K Seamans, and Terrence J Sejnowski. Neurocomputational models of working memory. *Nature neuroscience*, 3:1184–1191, 2000.

[22] Hermann Ebbinghaus. *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*. Duncker & Humblot, 1885.

[23] K Anders Ericsson and Walter Kintsch. Long-term working memory. *Psychological review*, 102(2):211, 1995.

[24] Cristóbal Esteban, Danilo Schmidt, Denis Krompaß, and Volker Tresp. Predicting sequences of clinical events by using a personalized temporal latent embedding model. In *Proceedings of the IEEE International Conference on Healthcare Informatics*, 2015.

[25] Cristóbal Esteban, Yinchong Yang, Stephan Baier, Denis Krompaß, and Volker Tresp. Modeling the event-driven temporal evolution of knowledge graphs. Submitted, 2015.

[26] Michael J Frank, Bryan Loughry, and Randall C OReilly. Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cognitive, Affective, & Behavioral Neuroscience*, 1 (2):137–160, 2001.

[27] Michael S Gazzaniga. *The cognitive neurosciences*. MIT press, 2004.

[28] Alex Graves, Greg Wayne, and Ivo Danihelka. Neural turing machines. *arXiv preprint arXiv:1410.5401*, 2014.

[29] Daniel L Greenberg and Mieke Verfaellie. Interdependence of episodic and semantic memory: evidence from neuropsychology. *Journal of the International Neuropsychological society*, 16(05):748–753, 2010.

[30] Thomas L Griffiths, Mark Steyvers, and Alana Firl. Google and the mind predicting fluency with pagerank. *Psychological Science*, 18(12):1069–1076, 2007.

[31] Thomas L Griffiths, Mark Steyvers, and Joshua B Tenenbaum. Topics in semantic representation. *Psychological review*, 114(2):211, 2007.

[32] Geoffrey E Hinton. Implementing semantic networks in parallel hardware. In *Parallel models of associative memory*, pages 161–187. Erlbaum, 1981.

[33] Geoffrey E Hinton and James A Anderson. *Parallel Models of Associative Memory: Updated Edition*. Psychology Press, 2014.

[34] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[35] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[36] John Jonides, Richard L Lewis, Derek Evan Nee, Cindy A Lustig, Marc G Berman, and Katherine Sledge Moore. The mind and brain of short-term memory. *Annual review of psychology*, 59:193, 2008.

[37] Charles Kemp, Joshua B. Tenenbaum, Thomas L. Griffiths, Takeshi Yamada, and Naonori Ueda. Learning systems of concepts with an infinite relational model. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*, volume 3 of *AAAI'06*, page 5, 2006.

[38] Teuvo Kohonen. *Self-organization and associative memory*, volume 8. Springer, 2012.

[39] Denis Krompaß, Maximilian Nickel, Xueyan Jiang, and Volker Tresp. Non-negative tensor factorization with rescal. *Tensor Methods for Machine Learning, ECML workshop*, 2013.

[40] Denis Krompaß, Xueyan Jiang, Maximilian Nickel, and Volker Tresp. Probabilistic Latent-Factor Database Models. In *Proceedings of the 1st Workshop on Linked Data for Knowledge Discovery (ECML PKDD)*, 2014.

[41] Denis Krompaß, Stephan Baier, and Volker Tresp. Type-constrained representation learning in knowledge graphs. In *The Semantic Web–ISWC 2015*, pages 640–655. Springer International Publishing, 2015.

[42] Thomas K Landauer and Susan T Dumais. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2):211, 1997.

[43] Thomas K Landauer, Peter W Foltz, and Darrell Laham. An introduction to latent semantic analysis. *Discourse processes*, 25(2-3):259–284, 1998.

[44] Kevin Lund, Curt Burgess, and Ruth Ann Atchley. Semantic and associative priming in high-dimensional semantic space. *Proceedings of the 17th annual conference of the Cognitive Science Society*, 17:660–665, 1995.

[45] James L McClelland and David E Rumelhart. Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, 114(2):159, 1985.

[46] Morten Mørup, Lars Kai Hansen, and Sidse M Arnfred. Algorithms for sparse nonnegative tucker decompositions. *Neural Computation*, 20(8):2112–2131, 2008.

[47] Maximilian Nickel. *Tensor factorization for relational learning*. PhD thesis, Ludwig Maximilian University of Munich, 2013.

[48] Maximilian Nickel and Volker Tresp. Learning Taxonomies from Multi-Relational Data via Hierarchical Link-Based Clustering. In *Learning Semantics. Workshop at NIPS'11*, Granada, Spain, 2011. URL `http://learningsemanticsnips2011.wordpress.com/`.

[49] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A Three-Way Model for Collective Learning on Multi-Relational Data. In *Proceedings of the 28th International Conference on Machine Learning*, 2011.

[50] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. Factorizing YAGO: scalable machine learning for linked data. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, pages 271–280, 2012. ISBN 978-1-4503-1229-5. doi: 10.1145/2187836.2187874. URL `http://doi.acm.org/10.1145/2187836.2187874`.

[51] Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. A review of relational machine learning for knowledge graphs: From multi-relational link prediction to automated knowledge graph construction. *Proceedings of the IEEE*, 2015.

[52] Maximilian Nickel, Lorenzo Rosasco, and Tomaso Poggio. Holographic embeddings of knowledge graphs. *arXiv preprint arXiv:1510.04935*, 2015.

[53] Randall C O'Reilly and Michael J Frank. Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural computation*, 18(2):283–328, 2006.

[54] Jeroen GW Raaijmakers and Richard M Shiffrin. SAM: A theory of probabilistic search of associative memory. *The psychology of learning and motivation: Advances in research and theory*, 14:207–262, 1981.

[55] Daniel L Schacter. Implicit memory: History and current status. *Journal of experimental psychology: learning, memory, and cognition*, 13(3):501, 1987.

[56] Lokendra Shastri. Episodic memory and cortico–hippocampal interactions. *Trends in cognitive sciences*, 6(4):162–168, 2002.

[57] Amit Singhal. Introducing the Knowledge Graph: things, not strings, May 2012. URL `http://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html`.

[58] Edward E Smith and Stephen M Kosslyn. *Cognitive Psychology: Pearson New International Edition: Mind and Brain*. Pearson Higher Ed, 2013.

[59] Larry R Squire. Memory and brain. *New York*, 1987.

[60] Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: A Core of Semantic Knowledge. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07. ACM, 2007.

[61] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lars Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1701–1708. IEEE, 2014.

[62] Endel Tulving. Episodic and semantic memory 1. *Organization of Memory. London: Academic*, 381 (e402):4, 1972.

[63] Endel Tulving. *Elements of episodic memory*. Oxford University Press, 1985.

[64] Endel Tulving. Episodic memory: from mind to brain. *Annual review of psychology*, 53(1):1–25, 2002.

[65] Jason Weston, Sumit Chopra, and Antoine Bordes. Memory networks. *arXiv preprint arXiv:1410.3916*, 2014.

[66] Zhao Xu, Volker Tresp, Kai Yu, and Hans-Peter Kriegel. Infinite Hidden Relational Models. In *Proceedings of the 22nd International Conference on Uncertainity in Artificial Intelligence*, pages 544–551, 2006.

[67] Eiling Yee, Evangelia G Chrysikou, and Sharon L Thompson-Schill. *The Cognitive Neuroscience of Semantic Memory*. Oxford Handbook of Cognitive Neuroscience, Oxford University Press, 2014.