Learning with Knowledge Graphs

Volker Tresp[1,2], Yunpu Ma[1,2], Stephan Baier[2]

[1] Siemens AG, Corporate Technology, Munich, Germany
[2] Ludwig-Maximilians-Universität München, Munich, Germany

Abstract. In recent years a number of large-scale triple-oriented knowledge graphs have been generated. They are being used in research and in applications to support search, text understanding and question answering. Knowledge graphs pose new challenges for machine learning, and research groups have developed novel statistical models that can be used to compress knowledge graphs, to derive implicit facts, and to detect errors in the knowledge graph. In this paper we decribe the concept of triple-oriented knowledge graphs and corresponding learning approaches. We also discuss episodic knowledge graphs which are able to represent temporal data; learning with episodic data can be the basis for decision support systems, e.g. in a clinical context. Finally we discuss how knowledge graphs can support perception, by mapping subsymbolic sensory inputs, such as images, to semantic triples. A particular feature of our approach would be that perception, episodic memory and semantic memory are highly interconnected and that, in a cognitive interpretation, all rely on the same brain structures.

1   Semantic Knowledge Graphs

A technical realization of a semantic memory is a knowledge graph (KG) which is a triple-oriented knowledge representation: A labelled link implies a *(subject, predicate, object)* statement where subject and object are entities that are represented as the nodes in the graph and where the predicate labels the link from subject to object. Large KGs have been developed that support search, text understanding and question answering [8]. A KG can be represented as a tensor which maps indices to true or false

$$s, p, o \mapsto Q$$

with $Q \in \{\mathtt{T}, \mathtt{F}\}$, and where $s \in 1, \ldots, N$ and $o \in 1, \ldots, N$ are indices for the N entities used as subject and object, and where $p \in 1, \ldots, R$ is the index for the predicate.

A statistical model for a KG can be obtained by a tensor model of the form

$$s, p, o \mapsto \mathbf{a}_{e(s)}, \mathbf{a}_p, \mathbf{a}_{e(o)} \mapsto P. \tag{1}$$

Here $e(s)$ and $e(o)$ are the entities associated with subject and object, respectively. The indices are first mapped to their latent representations $\mathbf{a}_{e(s)}, \mathbf{a}_p, \mathbf{a}_{e(o)}$

which are then mapped to a probability $P \in [0,1]$. $P((s,p,o) = \mathtt{T}|\mathbf{a}_{e(s)}, \mathbf{a}_p, \mathbf{a}_{e(o)})$ represents the Bernoulli probability that the triple (s,p,o) is true, and, when normalized across all triples, $P(s,p,o|\mathbf{a}_{e(s)}, \mathbf{a}_p, \mathbf{a}_{e(o)})$ stands for the categorical probability that the triple (s,p,o) is selected as an answer in a query process. A number of mathematical models have been developed for the mapping in Equation 1 (see [7]). A representative example is the RESCAL model [6], which is a constraint Tucker2 tensor model.

2 Episodic Knowledge Graphs

Whereas a semantic 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. Generalizing the semantic KG, an episodic KG can be represented as a 4-way tensor with time index t as the map

$$s, p, o, t \mapsto Q.$$

A statistical model for a KG can be obtained by a 4-way tensor model of the form

$$s, p, o, t \mapsto \mathbf{a}_{e(s)}, \mathbf{a}_p, \mathbf{a}_{e(o)}, \mathbf{a}_t \mapsto P \tag{2}$$

where $\mathbf{a}_t$ is the latent representation for time index t.

The basis for the tight link between different memory functions is the "unique representation hypothesis", which states that an entity has a unique latent representation in a technical application, but maybe also in the human brain [9].

As discussed in [11, 5] both the episodic KG and the semantic KG might rely on the same representations, i.e., it was proposed that the semantic KG can be derived from the episodic KG by a marginalization operation. Thus an episodic fact might represent that "Jack, wasDiagnosed, Diabetes, on Jan 15", the derived semantic fact might be "Jack, hasDisease, Diabetes". In [3, 4] medical decision systems are described that combine semantic and episodic tensor representations of data with recurrent neural network predictive models.

3 Perception

The tensor models permit generalization, i.e., the prediction of the probability of triples which were not known to be true in the data. This is especially important in perception, which we propose can be thought off as the mapping of subsymbolic sensory inputs to a semantic description in the form of a set of triples, describing and explaining the sensory inputs. These triples then becomes part of episodic memory.

Let $u_{t,1}, \ldots, u_{t,c}, \ldots, u_{t,C}$ be the content of the sensory buffers at time t. We propose that this sensory input can predict the latent representation for time in the form of a map

$$u_{t,1}, \ldots, u_{t,c}, \ldots, u_{t,C} \mapsto \mathbf{a}_t.$$

This map $\mathbf{a}_t(\mathbf{u}_t, \mathbf{w})$ might be modelled by a deep neural network with weights $\mathbf{w}$. Perceptual decoding then produces likely triples from the probability distribution (generalized nonlinear model) using

$$P(s, p, o; \mathbf{a}_{e(s)}, \mathbf{a}_p, \mathbf{a}_{e(o)}, \mathbf{a}_t(\mathbf{u}_t, \mathbf{w})).$$

An episodic memory would simply store $\mathbf{a}_t$, and memorizing simply means the restoring of a past $\mathbf{a}_t$, which then can be decoded as described [9, 10]. A semantic memory uses the marginalizing approach describes in Section 2.

As another approach, there is the option to use $P(s, p, o)$ or $P(s, p, o, t)$ as a semantic prior in sensory decoding. This was the basis for approaches to extract triples from Web sources [2] and for the extraction of triples from images [1].

References

[1] Stephan Baier, Yunpu Ma, and Volker Tresp. Improving visual relationship detection using semantic modeling of scene descriptions. In *ISWC*, 2017.

[2] 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 *KDD*, 2014.

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

[4] Yinchong Jang, Volker Tresp, and Peter Fasching. Predictive modeling of therapy decisions in metastatic breast cancer with recurrent neural network encoder and multinomial hierarchical regression decoder. In *ICHI*, 2017.

[5] Yunpu Ma, Volker Tresp, and Erik Daxberger. Embedding models for episodic memory. In *submitted*, 2017.

[6] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A Three-Way Model for Collective Learning. In *ICML*, 2011.

[7] 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.

[8] Amit Singhal. Introducing the Knowledge Graph: things, not strings. *Official Google Blog*, 2012.

[9] Volker Tresp, Cristóbal Esteban, Yinchong Yang, Stephan Baier, and Denis Krompaß. Learning with memory embeddings. *NIPS 2015 Workshop (extended TR); arXiv:1511.07972*, 2015.

[10] Volker Tresp, Yunpu Ma, and Stephan Baier. Tensor memories. In *CCN*, 2017.

[11] Volker Tresp, Yunpu Ma, Stephan Baier, and Yinchong Yang. Embedding learning for declarative memories. In *ESWC*, 2017.