Filling Gaps in Industrial Knowledge Graphs via Event-Enhanced Embedding

Martin Ringsquandl[1,2], Evgeny Kharlamov[3], Daria Stepanova[4], Marcel Hildebrandt[1], Steffen Lamparter[2], Raffaello Lepratti[5], Ian Horrocks[3], and Peer Kröger[1]

[1] Ludwig-Maximilians University [2] Siemens AG CT [3] University of Oxford
[4] Max-Planck Institut für Informatik [5] Digital Factory, Siemens PLM Software

Motivation. Knowledge Graphs (KGs) nowadays power many important applications including Web search[1], question answering [1], machine learning [13], data integration [7], entity disambiguation and linking [3, 5]. A KG is typically defined as a collection of triples $\langle entity, predicate, entity\rangle$ that form a directed graph where nodes are entities and edges are labeled with binary predicates (relations). Examples of large-scale KGs range from general-purpose such as Yago [17] and DBPedia [9] to domain specific ones such as Siemens [7] and Statoil [6] corporate KGs.

Large-scale KGs are often automatically constructed and highly incomplete [4] in the sense that they are missing certain triples. Due to their size and the speed of growth, manual completion of such KGs is infeasible. In order to address this issue, a number of relational learning approaches for *automatic KG completion* have been recently proposed, see [4, 10] for an overview. Many of these approaches are based on learning representations, or *embeddings*, of entities and relations [2, 11, 16]. It was shown that the quality of embeddings can be significantly improved if the embedding's vector space is enriched with additional information from an *external source*, such as a corpora of natural language text [19] or structural knowledge such as rules [18] or type constraints [8].

An important type of external knowledge that is common in practice and to the best of our knowledge has not been explicitly considered so far is *event log* data. Events naturally appear in many applications including social networks, smart cities, and manufacturing. In social networks the nodes of a KG can be people and locations, and edges can be friendship relations and places of birth [20], while an event log for a person can be a sequence of (possibly repetitive) places that the person has visited. In smart cities a KG can model traffic [15] by representing cameras, traffic lights, and road topology, while an event log for one day can be a sequence of traffic signals where jams or accidents have occurred. In smart manufacturing an event log can be a sequence of possible states, e.g., overheating or low power of machines such as conveyors, and these logs can be emitted during a manufacturing process.

In this work we define an event log for a KG as a set of sequences constituted of entities (possibly with repetitions) that may occur in the KG as nodes. Moreover, we assume that not every entity from a KG, but only what we call *event entities* can occur in logs. In the above: visited cities, traffic signals, and alarms are event entities. As we see later in the paper this separation of entities in a KG into event and non-event is important and practically motivated. We now illustrate an industrial KG and event log.

[1] `https://en.wikipedia.org/wiki/Knowledge_Graph`

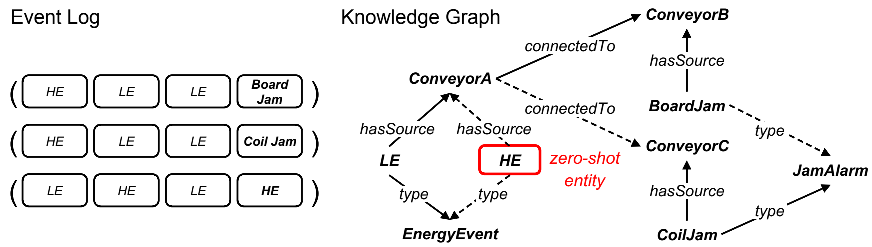

Fig. 1: Excerpt of a manufacturing KG and an event log.

Illustration of Scenarios. Consider an industrial KG that is inspired by a Siemens automated factory, that we will use later on for experiments, and that contains information about factory equipment, products, as well as materials and processes to produce the products. The KG was semi-automatically generated by parsing heterogeneous spreadsheets and other semi-structured data repositories and it is incomplete. In Figure 1 we depict a small excerpt from this KG where solid lines denote relations that are in the KG while dashed – the missing relations. The KG contains the topology of the conveyors A, B, and C and says that two of them (A and B) are connected to each other: $\langle ConveyorA, connectedTo, ConveyorB\rangle$. The KG also stores operator control specifications, in particular, event entities that the equipment can emit during operation. For example, $CoilJam$, is an event entity and it can be emitted by conveyor C, i.e., $\langle CoilJam, hasSource, ConveyorC\rangle$. Event entities have further semantics described by the typing, e.g. $CoilJam$ is of type $JamAlarm$, severity levels, and possible emitting source locations. At the same time, the KG misses the facts that the conveyors A and C are connected in the factory; that $BoardJam$ is of type $JamAlarm$, and $HighEnergy$ (HE) has the source $ConveyorA$ and is of type $EnergyEvent$.

Additionally, in the example, we assume that an event log recorded during the operation of the factory consists of three following sequences over event entities:

$$(HE, LE, LE, BoardJam), (HE, LE, LE, CoilJam), (LE, HE, LE, HE).$$

Observe that the event log suggests that a jam typically occurs after a sequence of two consecutive low energy consumption (LE) events.

Problem Statement. An event log gives external knowledge to the KG by specifying frequent sequential patterns on the KG's entities. These patterns capture some processes that the nodes of a KG can be involved in, i.e., manufacturing with machines described by the KG, traveling by a person mentioned in the KG, or traffic around traffic signals. This type of external knowledge has conceptual differences from text corpora where KG entities are typically described in a natural language and where occurrences of KG entities do not necessarily correspond to any process. Events are also different from rules or constraints that introduce formal restrictions on some relations.

The goal of this work is to understand how event logs can enhance relational learning for KGs. We address this problem by proposing an *Event-enhanced Knowledge Learning* (*EKL*) approach for KG completion that intuitively has two sub–steps:

1. *Event alignment*, where event entities are aligned in a low-dimensional vector space that reflects sequential similarity, and
2. *KG completion*, where the KG is extended with missing edges that can be either event-specific, e.g., such as the *type* edge between $BoardJam$ and $JamAlarm$

in our rubbing example, or not event-specific, e.g., such as *connectedTo* between $ConveyorA$ and $ConveyorC$ in the running example.

Observe, the event logs *directly* influence the first step while also *indirectly* the second step of EKL. Hence, we expect a collective learning effect in a sense that the overall KG completion can benefit from event alignment, and vice versa.

Illustration of Ideas. During the first step EKL will align $BoardJam$ and $CoilJam$ to be similar. In the second step EKL will accordingly adjust entities $ConveyorC$ and $ConveyorB$ and then predict that $ConveyorA$ is likely to also be connected to $ConveyorC$. Intuitively the missing link between the conveyors can be inferred from the sequential pattern in the event log: the log tells us that both $BoardJam$ and $CoilJam$ occur as a consequence of two consecutive LE events and therefore exhibit similar semantics. This similarity is carried to conveyor entities B and C, which leads to an increased likelihood that they both follow the same entity $ConveyorA$.

Note that the prediction of event-specific missing links is not the standard task for relational learning since we are predicting links *within* the background. Moreover, our approach can address the *zero-shot scenario*, where some event entities only appear in the event log, but they are novel to the KG (it is marked with red in Figure 1). E.g., HE in the running example corresponds to an entity that is missing in the KG, that has to be aligned during the first step of EKL and then linked to $ConveyorA$ as well as to its type during the second step of EKL. Thus, EKL can also populate a KG with new (unseen) entities.

Contributions. The contributions of our work are as follows:

- We proposed several EKL approaches to KG completion that comprise
  - two model architectures that allow to combine (representations of) a KG and an event log for simultaneous training of both representations; this requires a non-trivial design of a model architecture that reflects interconnections of shared embeddings,
  - three models for event logs that reflect different notions of event context.
- We conducted an extensive evaluation of our approach and comparison to a state-of-the-art baseline on real-world data from a factory, on smart city traffic data, and controlled experiment data. Our results show that we significantly outperform two state-of-the-art baselines and the advantages are most visible for predicting links between entities that reflect the sequential process nature within the KG.

We presented a very preliminary version of this work as a short in-use paper [14] and a longer version as a research paper [12].

Acknowledgements. This work was partially supported by the EPSRC projects DBOnto, MaSI^3 and ED^3.

References

1. Bordes, A., Chopra, S., Weston, J.: Question answering with subgraph embeddings. In: EMNLP. pp. 615–620 (2014)
2. Bordes, A., Usunier, N., García-Durán, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: NIPS. pp. 2787–2795 (2013)
3. Cucerzan, S.: Large-scale named entity disambiguation based on wikipedia data. In: EMNLP-CoNLL. pp. 708–716 (2007)
4. Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmann, T., Sun, S., Zhang, W.: Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In: ACM SIGKDD. pp. 601–610 (2014)
5. Hachey, B., Radford, W., Nothman, J., Honnibal, M., Curran, J.R.: Evaluating entity linking with wikipedia. Artif. Intell. 194, 130–150 (2013)
6. Kharlamov, E., Hovland, D., Skjæveland, M.G., Bilidas, D., Jiménez-Ruiz, E., Xiao, G., Soylu, A., Lanti, D., Rezk, M., Zheleznyakov, D., Giese, M., Lie, H., Ioannidis, Y.E., Kotidis, Y., Koubarakis, M., Waaler, A.: Ontology based data access in statoil. JWS 44, 3–36 (2017)
7. Kharlamov, E., Mailis, T., Mehdi, G., Neuenstadt, C., Özçep, Ö.L., Roshchin, M., Solomakhina, N., Soylu, A., Svingos, C., Brandt, S., Giese, M., Ioannidis, Y.E., Lamparter, S., Möller, R., Kotidis, Y., Waaler, A.: Semantic access to streaming and static data at siemens. JWS 44, 54–74 (2017)
8. Krompaß, D., Baier, S., Tresp, V.: Type-Constrained Representation Learning in Knowledge Graphs. ISWC (2015)
9. Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., Bizer, C.: Dbpedia - A large-scale, multilingual knowledge base extracted from wikipedia. Semantic Web 6(2), 167–195 (2015)
10. Nickel, M., Murphy, K., Tresp, V., Gabrilovich, E.: A review of relational machine learning for knowledge graphs. Proceedings of the IEEE 104(1), 11–33 (2016)
11. Nickel, M., Rosasco, L., Poggio, T.A.: Holographic embeddings of knowledge graphs. In: AAAI. pp. 1955–1961 (2016)
12. Ringsquandl, M., Kharlamov, E., Stepanova, D., Hildebrandt, M., Lamparter, S., Lepratti, R., Horrocks, I., Kroeger, P.: Event-enhanced learning for knowledge graph completion. In: ESWC (2018)
13. Ringsquandl, M., Lamparter, S., Brandt, S.P., Lepratti, R.: Semantic-guided Feature Selection for Industrial Automation Systems. In: ISWC. Springer (2015)
14. Ringsquandl, M., Lamparter, S., Kharlamov, E., Lepratti, R., Stepanova, D., Kroeger, P., Horrocks, I.: On event-driven learning of knowledge in smart factories: The case of siemens. In: IEEE Big Data (2017)
15. Santos, H., Dantas, V., Furtado, V., Pinheiro, P., McGuinness, D.L.: From data to city indicators: A knowledge graph for supporting automatic generation of dashboards. In: ESWC. pp. 94–108 (2017)
16. Shi, B., Weninger, T.: ProjE : Embedding Projection for Knowledge Graph Completion. AAAI 2017 pp. 1–14 (2017)
17. Suchanek, F.M., Kasneci, G., Weikum, G.: Yago: a core of semantic knowledge. In: Proc. of WWW. pp. 697–706 (2007)
18. Wang, Q., Wang, B., Guo, L.: Knowledge base completion using embeddings and rules. In: IJCAI. pp. 1859–1866 (2015)
19. Wang, Z., Li, J.Z.J.: Text-Enhanced Representation Learning for Knowledge Graph. IJCAI pp. 1293–1299 (2016)
20. Yang, Z., Tang, J., Cohen, W.W.: Multi-modal bayesian embeddings for learning social knowledge graphs. In: IJCAI. pp. 2287–2293 (2016)