Knowledge Graph Representation Learning and NLP Applications

Shaoxiong Ji

shaoxiong.ji@aalto.fi

Research Club at Silo.AI
(14/04/2022)

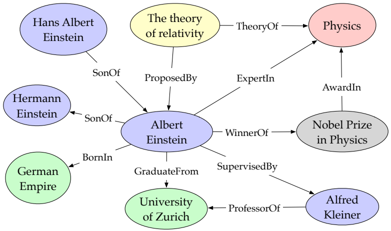
Outline

Introduction

A knowledge graph is a structured representation of facts, consisting of entities, relationships, and semantic descriptions.

(Albert Einstein, **BornIn**, German Empire)
(Albert Einstein, **SonOf**, Hermann Einstein)
(Albert Einstein, **GraduateFrom**, University of Zurich)
(Albert Einstein, **WinnerOf**, Nobel Prize in Physics)
(Albert Einstein, **ExpertIn**, Physics)
(Nobel Prize in Physics, **AwardIn**, Physics)
(The theory of relativity, **TheoryOf**, Physics)
(Albert Einstein, **SupervisedBy**, Alfred Kleiner)
(Alfred Kleiner, **ProfessorOf**, University of Zurich)
(The theory of relativity, **ProposedBy**, Albert Einstein)
(Hans Albert Einstein, **SonOf**, Albert Einstein)

(a) Factual triples in knowledge base.



(b) Entities and relations in knowledge graph.

Figure: An example of knowledge base and knowledge graph.

(Ji et al., 2022)

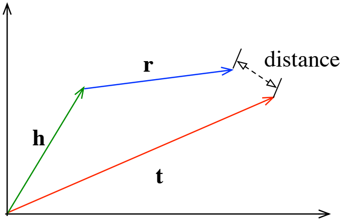
Why Knowledge Graphs?

- Context:
  - entities (real-world objects and abstract concepts)
  - relation between entities, types, and properties
- Relational Reasoning
- Intrinsic Explainability
- Others: efficient search, flexibility ...

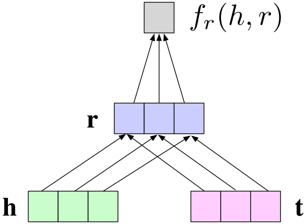
This presentation:

- How to represent and encode factual knowledge triples
- Knowledge graphs applied to NLP and healthcare tasks.

How to Learn Knowledge Graph Representation



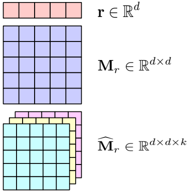
(a) Translational distance-based scoring of TransE.



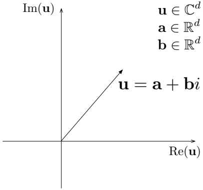
(b) Semantic similarity-based scoring of DistMult.

Figure: Illustrations of distance-based and similarity matching based scoring functions taking TransE (Bordes et al., 2013) and DistMult (Yang et al., 2015) as examples.

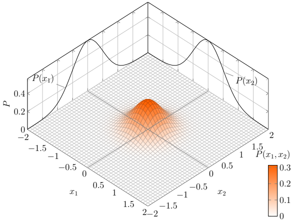
Representation Space



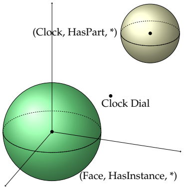
(a) Point-wise space.



(b) Complex vector space.



(c) Gaussian distribution.



(d) Manifold space.

- Point-Wise Space: widely applied
- Complex Space: real and imaginary parts
- Gaussian distribution: (un)certainties of entities and relations
- Manifold Space: a set of points with neighborhoods by the set theory

Encoding Models

encode the interactions of entities and relations through specific model architectures,

- linear/bilinear models: relations as a linear/bilinear mapping by projecting head entities into a representation space close to tail entities
- factorization models: decompose relational data into low-rank matrices for representation learning
- neural networks: encode relational data with non-linear neural activation and more complex network structures by matching semantic similarity of entities and relations

Linear/Bilinear and Factorization Models

Linear scoring function

$$g_r(\mathrm{h}, \mathrm{t}) = \mathrm{M}_r^T \begin{pmatrix} \mathrm{h} \\ \mathrm{t} \end{pmatrix} \tag{1}$$

Bilinear scoring function

$$f_r(h, t) = \mathrm{h}^\top \mathrm{M}_r \mathrm{t} \tag{2}$$

Factorization

RESCAL (Nickel et al., 2011)

three-way rank-r factorization RESCAL over each relational slice of knowledge graph tensor.

$$\mathcal{X}_k \approx \mathrm{A}\mathrm{R}_k\mathrm{A}^T \tag{3}$$

Neural Networks

MLP

$$f_r(h, t) = \sigma(\mathbf{w}^\top \sigma(\mathbf{W}[\mathbf{h}, \mathbf{r}, \mathbf{t}])), \quad (4)$$

NTN (Socher et al., 2013)

$$f_r(h, t) = \mathbf{r}^\top \sigma(\mathbf{h}^\top \widehat{\mathbf{M}}\mathbf{t} + \mathbf{M}_{r,1}\mathbf{h} + \mathbf{M}_{r,2}\mathbf{t} + \mathbf{b}_r), \quad (5)$$

CNN input triples into dense layer and convolution operation to learn semantic representation

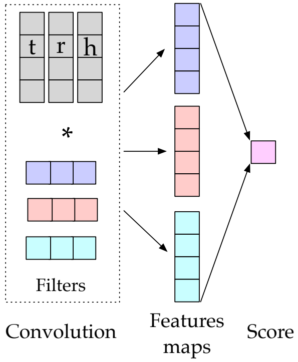


Figure: CNN (Nguyen et al., 2018)

Neural Networks

GCN (Shang et al., 2019) acts as encoder of knowledge graphs to produce entity and relation embeddings.

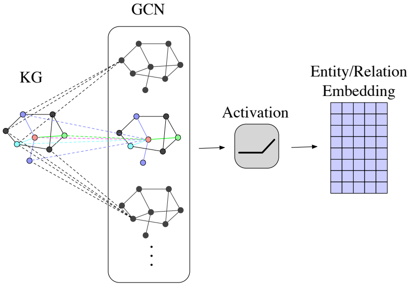


Figure: GCN (Shang et al., 2019)

Neural Networks

RNN

RSN encodes entity-relation sequences and skips relations discriminatively.

$$h'_t = \begin{cases} h_t & x_t \in \mathcal{E} \\ S_1 h_t + S_2 x_{t-1} & x_t \in \mathcal{R} \end{cases} \tag{6}$$

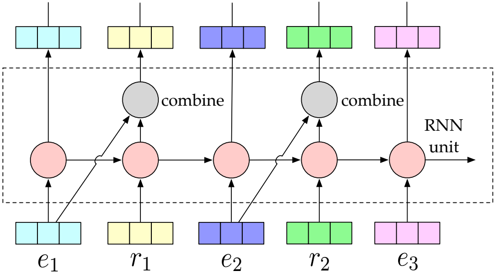


Figure: RSN (Guo et al., 2019)

Neural Networks

Transformers

Transformer-based CoKE (Wang et al., 2019) encodes triples as sequences with an entity replaced by `[MASK]`.

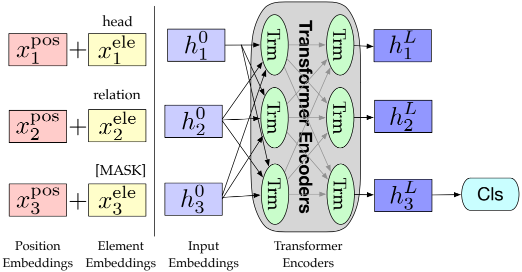


Figure: CoKE (Wang et al., 2019)

Incorporate Knowledge Graphs for NLP

Three paradigms:

- Incorporate large-scale knowledge graph in pretraining (in Sec. 2)
- Inject knowledge into model architecture [e.g., SentiLSTM (Ma et al., 2018)]
- Infuse knowledge-aware representations into text features [e.g., CompareNet (Hu et al., 2021)]

Knowledge-aware Language Models

- The success gained by pretraining self-supervised language models
- Integrating factual knowledge into language representation via pretraining

Knowledge-aware Language Models

Discriminating informative entities

- ERNIE-Tsinghua (Zhang et al., 2019): entity & random masking
- ERNIE-Baidu (Sun et al., 2019): entity and phrase masking
- GLM (Shen et al., 2020): graph-guided (linking) entity masking

Subgraph extraction

- BERT-MK (He et al., 2020): knowledge subgraph (multi-head attention)
- CoLAKE (Sun et al., 2020): word graph & knowledge subgraph (concat)

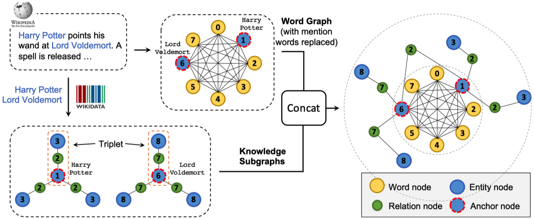
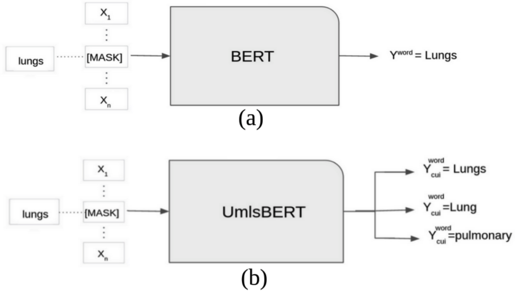


Figure: CoLAKE (Sun et al., 2020)

Knowledge-aware Language Models

Multi-label Concept (Synonym) Modeling

- UmlsBERT (Michalopoulos et al., 2021): concepts in UMLS Metathesaurus



Joint Training

- KEPLER (Wang et al., 2020): knowledge embedding + MLM losses

NLP Applications: Sentiment Analysis

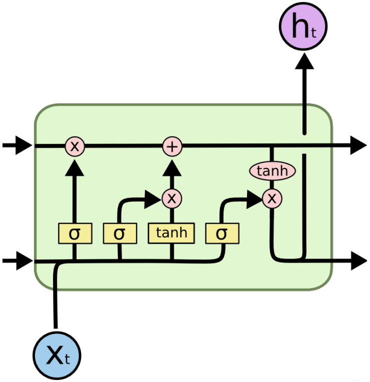


Figure: LSTM (Source: C. Olah)

Sentic LSTM (Ma et al., 2018)

$$
\begin{aligned}
f_i &= \sigma(W_f[x_i, h_{i-1}, \mu_i] + b_f) \\
I_i &= \sigma(W_I[x_i, h_{i-1}, \mu_i] + b_I) \\
\widetilde{C}_i &= tanh(W_C[x_i, h_{i-1}] + b_C) \\
C_i &= f_i * C_{i-1} + I_i * \widetilde{C}_i \\
o_i &= \sigma(W_o[x_i, h_{i-1}, \mu_i] + b_o) \\
o_i^c &= \sigma(W_{co}[x_i, h_{i-1}, \mu_i] + b_{co}) \\
h_i &= o_i * tanh(C_i) + o_i^c * tanh(W_c \mu_i)
\end{aligned}
$$

μ_i: knowledge concepts

NLP Applications: Fake News Detection

Gating

$$e_{\mathrm{KB}} = g_e \odot e_s + (1 - g_e) \odot e_d$$

Entity Comparison

$$a_i = f_{\mathrm{cmp}}\left(e_c, W_e \cdot e_{\mathrm{KB}}\right)$$

$$f_{\mathrm{cmp}}(x, y) = W_a[x - y, x \odot y]$$

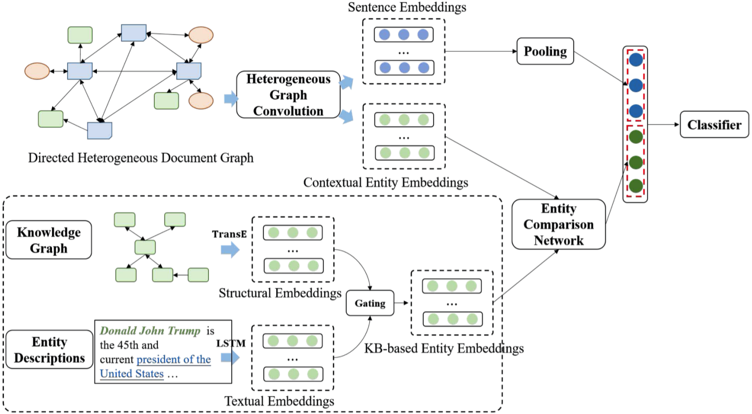


Figure: CompareNet (Hu et al., 2021)

NLP Applications: Healthcare

Medical NLI (Sharma et al., 2019)

$$e_w = e_{\text{BioELMo},w} \oplus e_{\text{DistMult},w} \oplus e_{\text{Senti}_w}$$

Medical Code Prediction

- Code-wise attention
  $\alpha_i = \text{softmax}\left(\mathrm{H}^\top \mathrm{v}_i\right)$
- Aggregation:
  $\mathrm{U} = \lambda \mathrm{W}_s \mathrm{S} + \mathrm{D}^\top \mathrm{W}_d$

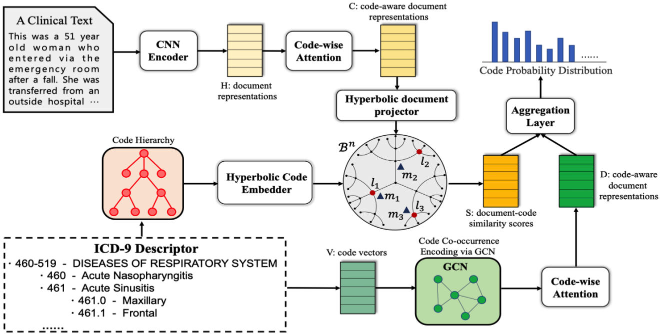


Figure: HyperCore (Cao et al., 2020)

Summary

Today's presentation:

- Knowledge Graph Representation Learning (three aspects)
- Knowledge-aware NLP Applications (three paradigms)

More applications of KGs:

- Ontology-guided distant supervision
- Healthcare: drug-drug interaction and drug adverse event detection
- And more, e.g., question answering and recommendation systems

Open Questions:

- Construction of knowledge graphs (manual v.s. automatic)
- Knowledge graph database indexing and query
- Dynamic knowledge graphs

References I

A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. In *NIPS*, pages 2787–2795, 2013.

P. Cao, Y. Chen, K. Liu, J. Zhao, S. Liu, and W. Chong. HyperCore: Hyperbolic and Co-graph Representation for Automatic ICD Coding. In *Proceedings of ACL*, pages 3105–3114, 2020.

L. Guo, Z. Sun, and W. Hu. Learning to exploit long-term relational dependencies in knowledge graphs. In *ICML*, pages 2505–2514, 2019.

B. He, D. Zhou, J. Xiao, Q. Liu, N. J. Yuan, T. Xu, et al. BERT-MK: Integrating Graph Contextualized Knowledge into Pre-trained Language Models. In *Findings of EMNLP*, page 2281–2290, 2020.

L. Hu, T. Yang, L. Zhang, W. Zhong, D. Tang, C. Shi, N. Duan, and M. Zhou. Compare to the knowledge: Graph neural fake news detection with external knowledge. In *Proceedings of ACL-IJCNLP*, pages 754–763, 2021.

S. Ji, S. Pan, E. Cambria, P. Marttinen, and P. S. Yu. A survey on knowledge graphs: Representation, acquisition and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 33, 2022.

Y. Ma, H. Peng, and E. Cambria. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm. In *AAAI*, pages 5876–5883, 2018.

G. Michalopoulos, Y. Wang, H. Kaka, H. Chen, and A. Wong. Umlsbert: Clinical domain knowledge augmentation of contextual embeddings using the unified medical language system metathesaurus. In *Proceedings of NAACL-HLT*, pages 1744–1753, 2021.

D. Q. Nguyen, T. D. Nguyen, D. Q. Nguyen, and D. Phung. A novel embedding model for knowledge base completion based on convolutional neural network. In *NAACL*, pages 327–333, 2018.

M. Nickel, V. Tresp, and H.-P. Kriegel. A three-way model for collective learning on multi-relational data. In *ICML*, volume 11, pages 809–816, 2011.

C. Shang, Y. Tang, J. Huang, J. Bi, X. He, and B. Zhou. End-to-end structure-aware convolutional networks for knowledge base completion. In *AAAI*, volume 33, pages 3060–3067, 2019.

S. Sharma, B. Santra, A. Jana, S. Tokala, N. Ganguly, and P. Goyal. Incorporating domain knowledge into medical nli using knowledge graphs. In *Proceedings of EMNLP-IJCNLP*, pages 6092–6097, 2019.

T. Shen, Y. Mao, P. He, G. Long, A. Trischler, and W. Chen. Exploiting structured knowledge in text via graph-guided representation learning. In *EMNLP*, 2020.

References II

R. Socher, D. Chen, C. D. Manning, and A. Ng. Reasoning with neural tensor networks for knowledge base completion. In *NIPS*, pages 926–934, 2013.

T. Sun, Y. Shao, X. Qiu, Q. Guo, Y. Hu, X.-J. Huang, and Z. Zhang. CoLAKE: Contextualized Language and Knowledge Embedding. In *COLING*, pages 3660–3670, 2020.

Y. Sun, S. Wang, Y. Li, S. Feng, X. Chen, H. Zhang, X. Tian, D. Zhu, H. Tian, and H. Wu. ERNIE: Enhanced representation through knowledge integration. *arXiv preprint arXiv:1904.09223*, 2019.

Q. Wang, P. Huang, H. Wang, S. Dai, W. Jiang, J. Liu, Y. Lyu, Y. Zhu, and H. Wu. CoKE: Contextualized knowledge graph embedding. *arXiv preprint arXiv:1911.02168*, 2019.

X. Wang, T. Gao, Z. Zhu, Z. Liu, J. Li, and J. Tang. KEPLER: A unified model for knowledge embedding and pre-trained language representation. *TACL*, 2020.

B. Yang, W.-t. Yih, X. He, J. Gao, and L. Deng. Embedding entities and relations for learning and inference in knowledge bases. In *ICLR*, pages 1–13, 2015.

Z. Zhang, X. Han, Z. Liu, X. Jiang, M. Sun, and Q. Liu. ERNIE: Enhanced language representation with informative entities. In *ACL*, pages 1441–1451, 2019.