FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

Karlsruher Institut für Technologie

Entity Type Prediction in Knowledge Graphs using Embeddings

Russa Biswas, Radina Sofronova, Mehwish Alam, and Harald Sack

02.06.2020

Motivation

What are the types of the following entities?

Violin

Instrument

Lisbon

City

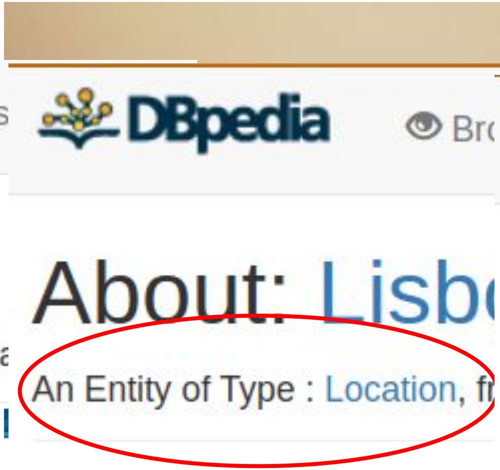
Yellow billed duck

Bird

Motivation

What are the types of the following entities?

Violin	Lisbon	Yellow billed duck



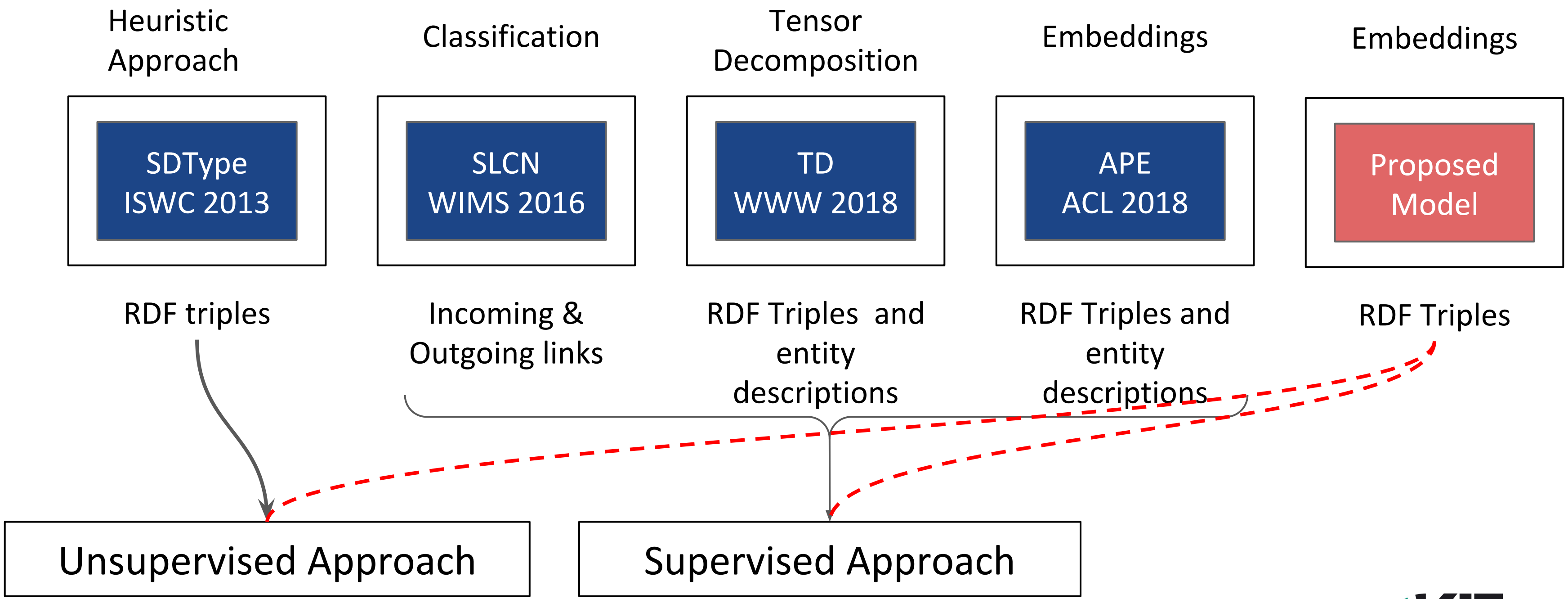
Instrument	City	Bird

Motivation

Coarse grained Type information in DBpedia at a glance

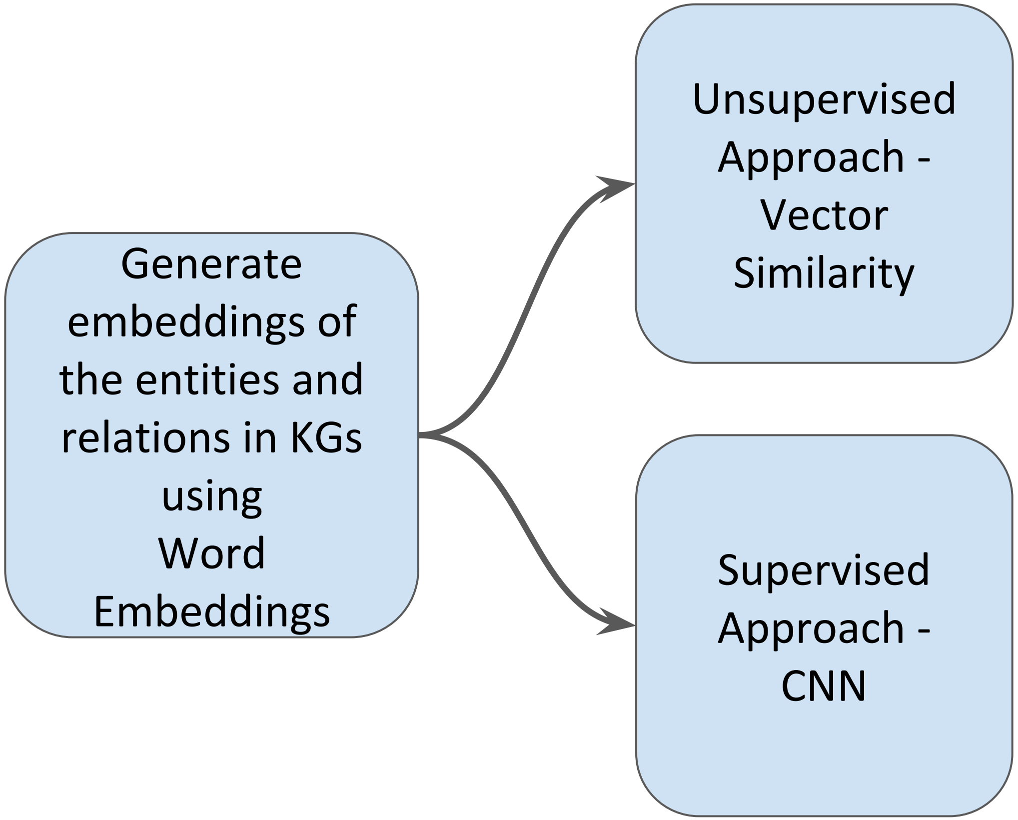
Classes	#Total entities	Percentage of entities with more fine grained type
dbo:Person	1,818,072	36.6%
dbo:Scientist	25,760	3.5%
dbo:Settlement	581,293	68.3%
dbo:Company	109,629	13.9%

Related Work



Russa Biswas et al. Entity Type Prediction in KGs using Embeddings. DL4KG Workshop @ ESWC 2020.

Proposed Approach



- Three different word embedding models are used to model the KGs
  - Word2Vec
  - FastText
  - GloVe
- Entity Typing is done based on these 3 word embedding models separately and are compared against each other.

Karlsruher Institut für Technologie
FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

Embeddings

Input: <dbr:Albert_Einstein, dbo:birthPlace, dbr:Ulm>.
<dbr:Albert_Einstein, dbo:field, dbr:Physics>.

Sentences

words

Embedding models are trained on triples with **Object Properties**

Word2Vec: Continuous Bag of Words (CBOW) approach is used.

FastText: Continuous Bag of Words (CBOW) approach is used.

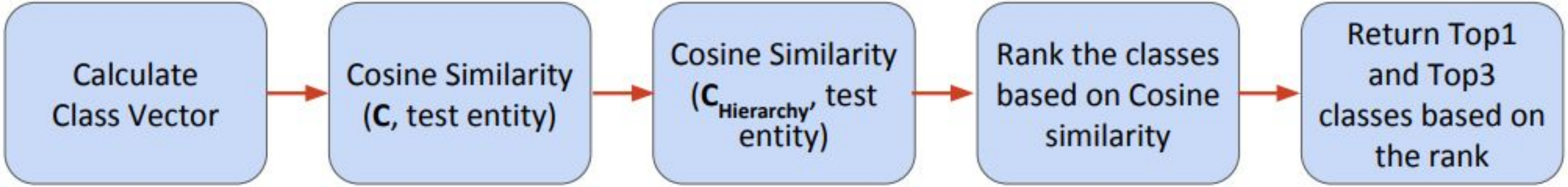
GloVe: Word co-occurrence matrix is used to learn the model.

Karlsruher Institut für Technologie

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

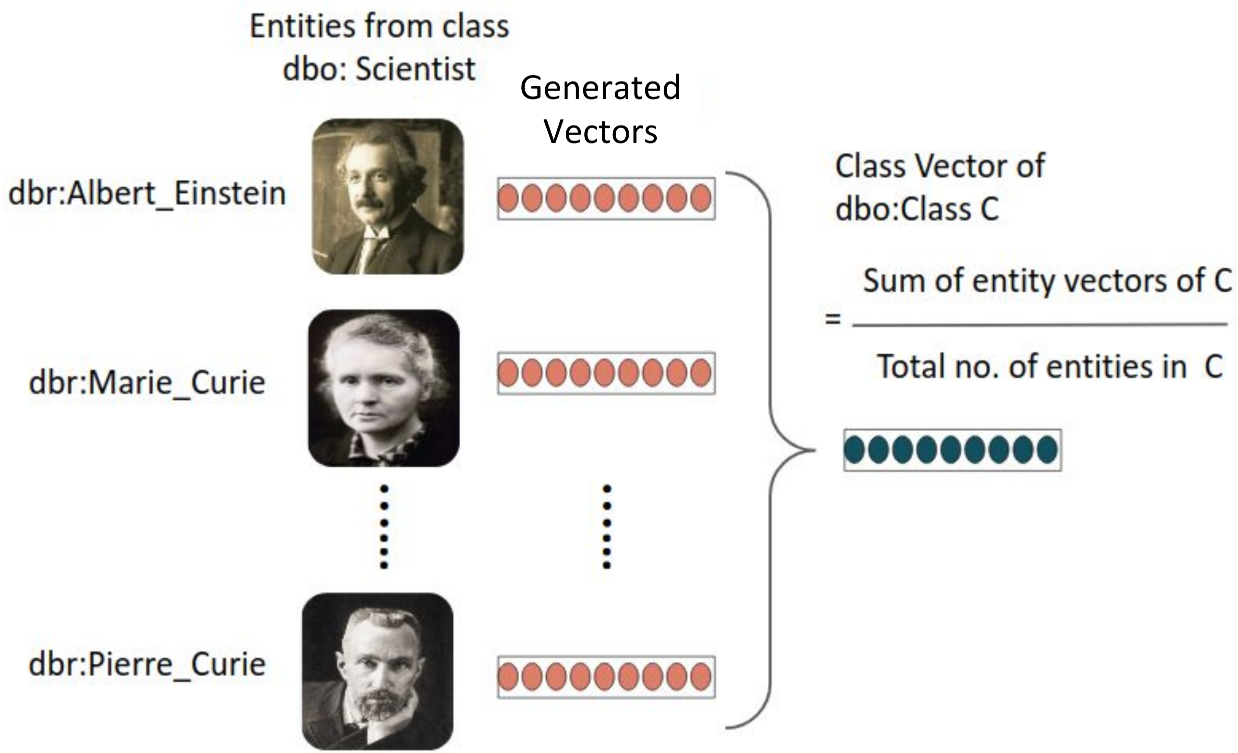
Pipeline of the Unsupervised Approach

Unsupervised approach is based on the vector similarity between the class vector and entity vector



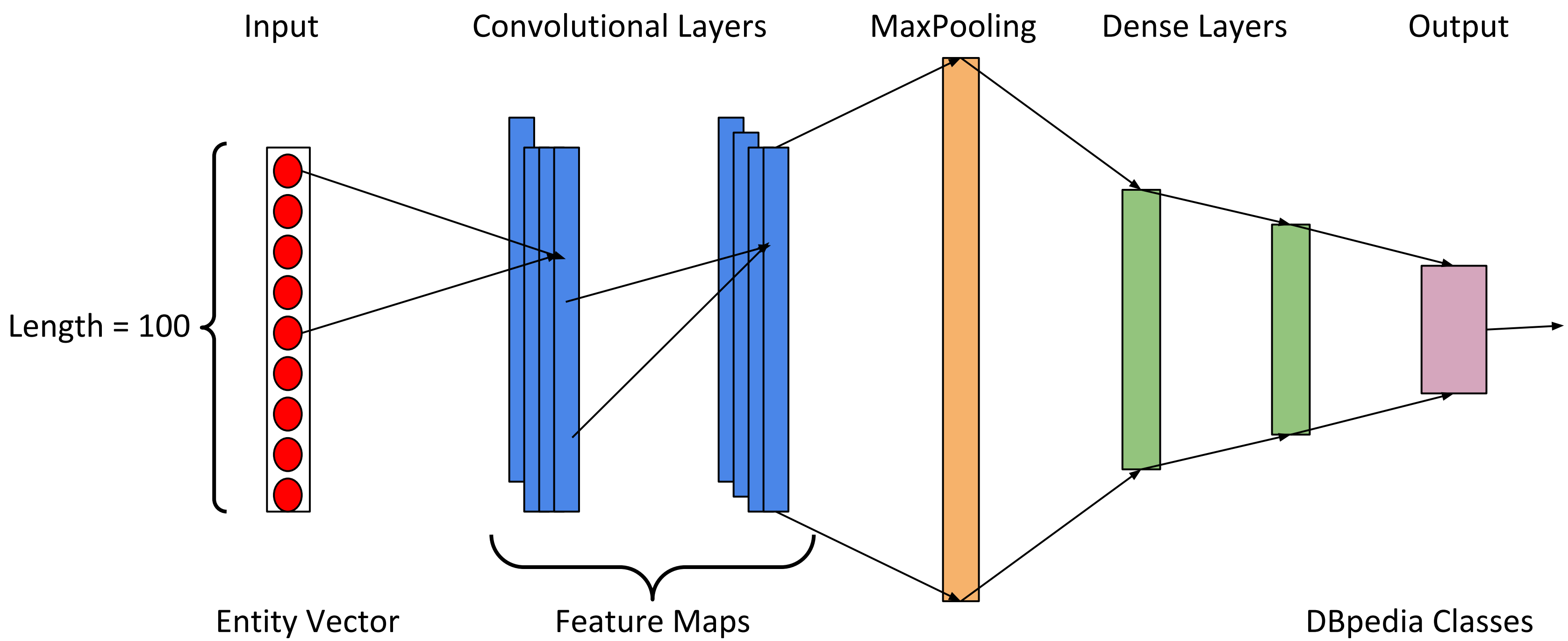
Russa Biswas et al. Entity Type Prediction in KGs using Embeddings. DL4KG Workshop @ ESWC 2020.

Generation of Class Vectors



Russa Biswas et al. Entity Type Prediction in KGs using Embeddings. DL4KG Workshop @ ESWC 2020.

Supervised Approach - 1D CNN



Russa Biswas et al. Entity Type Prediction in KGs using Embeddings. DL4KG Workshop @ ESWC 2020.

Karlsruher Institut für Technologie
FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

Datasets

1. **Dataset 1**:
   a. 59 less popular classes with the following characteristics:
      i. **15 classes** that have less than **500 entities per class**,
      ii. **20 classes** that have entities between **500 and 1000 entities per class**,
      iii. **24 classes** have more than **1000 entities per class**,
      iv. **Max.** no. of entities per class in this dataset is **500**, and
      v. **Min.** no. of entities per class in this dataset is **276**.
2. **Dataset 2**: **86** classes with **2k entities per class**.
3. **Dataset 3**: **81** classes with **4k entities per class**.

Results

Datasets	Models (Results in Percentage Accuracy)								
	Word2Vec			FastText			GloVe		
	Vector Similarity		CNN	Vector Similarity		CNN	Vector Similarity		CNN
	Hits@3	Hits@1		Hits@3	Hits@1		Hits@3	Hits@1	
Dataset 1	47.83	28.46	**56**	29.81	17.44	54	7.07	3.54	53.7
Dataset 2	58	39.4	**58.4**	43.81	31.16	56	15.9	8.2	55
Dataset 3	58	39.7	**62**	44.3	31.4	59	16.2	8.4	55.8

Russa Biswas et al. Entity Type Prediction in KGs using Embeddings. DL4KG Workshop @ ESWC 2020.

Results - Comparison with SDType

Test Dataset: The common entities between our dataset and the entities for which SDType model[1] predicts a change are considered.

Datasets	#Test Entities	SDType	Vector Similarity (Accuracy in Percentage)					
			Word2Vec		FastText		GloVe	
			Hits@1	Hits@3	Hits@1	Hits@3	Hits@1	Hits@3
Dataset 1	7425	83.35	32	63.78	12.58	25.87	2.8	11.04
Dataset 2	57467	80.43	46.94	69.53	38	61.8	16.37	30.4
Dataset 3	109948	81.22	48.21	71.6	39.54	64.14	17.07	31.58

http://wiki.dbpedia.org/services-resources/documentation/datasets#InstanceTypesSdtypedDbo

Conclusion and Future Work

- **Word embeddings** when applied on the Knowledge Graphs can be efficiently used for the task of Entity Type Prediction.

- **Word2Vec proves** to be the **best word embedding approach** out of the three word embedding approaches used **in KGs**.

- Supervised approach, **1D CNN works better than the unsupervised approach** for the task

- In Future Work, more information from the DBpedia such as **Datatype properties** are to be explored for the type prediction task.

References

Literature:

1. Melo A, Paulheim H, Völker J. Type prediction in RDF knowledge bases using hierarchical multilabel classification. InProceedings of the 6th International Conference on Web Intelligence, Mining and Semantics 2016 Jun 13 (p. 14). ACM.
2. Jin H, Hou L, Li J, Dong T. Attributed and Predictive Entity Embedding for Fine-Grained Entity Typing in Knowledge Bases. InProceedings of the 27th International Conference on Computational Linguistics 2018 Aug (pp. 282-292).
3. Paulheim H, Bizer C. Type inference on noisy rdf data. In International semantic web conference 2013 Oct 21 (pp. 510-525). Springer, Berlin, Heidelberg.
4. Moniruzzaman AB, Nayak R, Tang M, Balasubramaniam T. Fine-grained Type Inference in Knowledge Graphs via Probabilistic and Tensor Factorization Methods. In The World Wide Web Conference 2019 May 13 (pp. 3093-3100). ACM.
5. Pirrò G. Explaining and suggesting relatedness in knowledge graphs. In International Semantic Web Conference 2015 Oct 11 (pp. 622-639). Springer, Cham.

Images:

https://en.wikipedia.org/wiki/Lisbon#/media/File:Montagem_de_Lisboa.png
https://en.wikipedia.org/wiki/Violin#/media/File:Violin_VL100.png
https://en.wikipedia.org/wiki/Yellow-billed_duck#/media/File:Yellow-billed_Duck_Plettenbergbay_RWD.jpg

Karlsruher Institut für Technologie

FIZ Karlsruhe

Leibniz-Institut für Informationsinfrastruktur

15

Thank you

Homepage: https://www.fiz-karlsruhe.de/en/forschung/lebenslauf-und-publikationen-russa-biswas
Email id: russa.biswas@fiz-karlsruhe.de
Twitter: @russa.biswas

Karlsruher Institut für Technologie

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

Mitglied der
Leibniz
Leibniz Gemeinschaft