UNIVERSITY OF MISSOURI

Creation of EMA-KN – A Knowledge Network for Ecological Momentary Assessment

Cade Winters
Truman State

George Stafford, Justin Varghese, Fengxiang Zhao, Dr. Yi Shang
University of Missouri

INTRODUCTION

Domain-specific knowledge is necessary for critical analysis and decision-making in any scientific field. As a result, it is important that we have mechanisms for collecting and applying knowledge contributed by the larger scientific community. The current paradigm involves collecting knowledge in human-readable scientific papers across various scientific journals. The problem with this approach is that these papers are generally long and very dense, making the extraction of useful information a highly-specialized and labor-intensive task. It is for these reasons that we present EMA-KN, an automatically generated knowledge network pipeline built using the AI-KG framework.

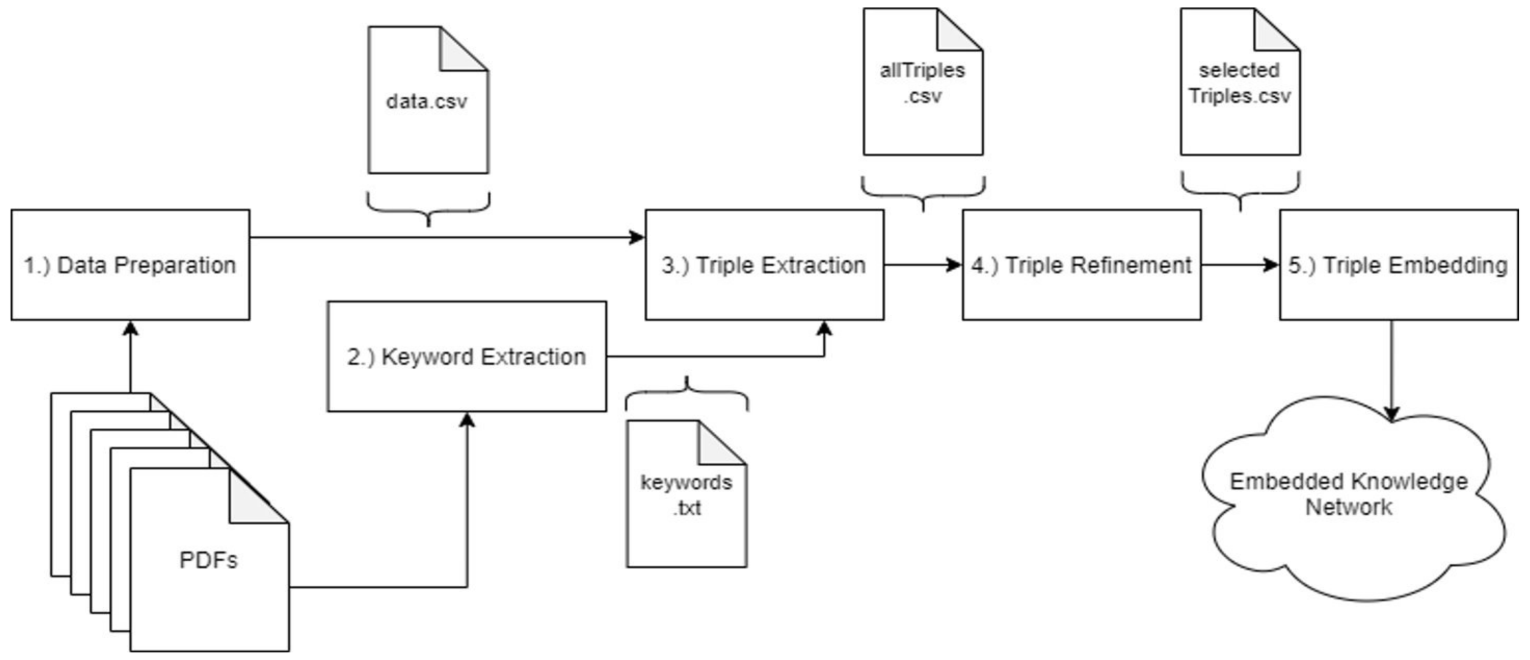


Figure 1: Use case diagram for knowledge network construction and embedding using the EMA-KN Pipeline

EXTRACTION

Our RDF Triple Extraction process makes use of the DyGIE++ extractor, and an OpenIE Extractor in order to properly extract the most beneficial RDF Triples from EMA related scientific papers. After our pipeline finishes extracting triples from the input dataset, it uses a combination of a list of generated keywords found from the input dataset and statistics in order to properly validate which RDF triples are considered to be golden(In that they convey true information about the world). The extraction process that we based ours on originally used a large ontology to base its data on, however we have altered it to no longer need a large and existing ontology.

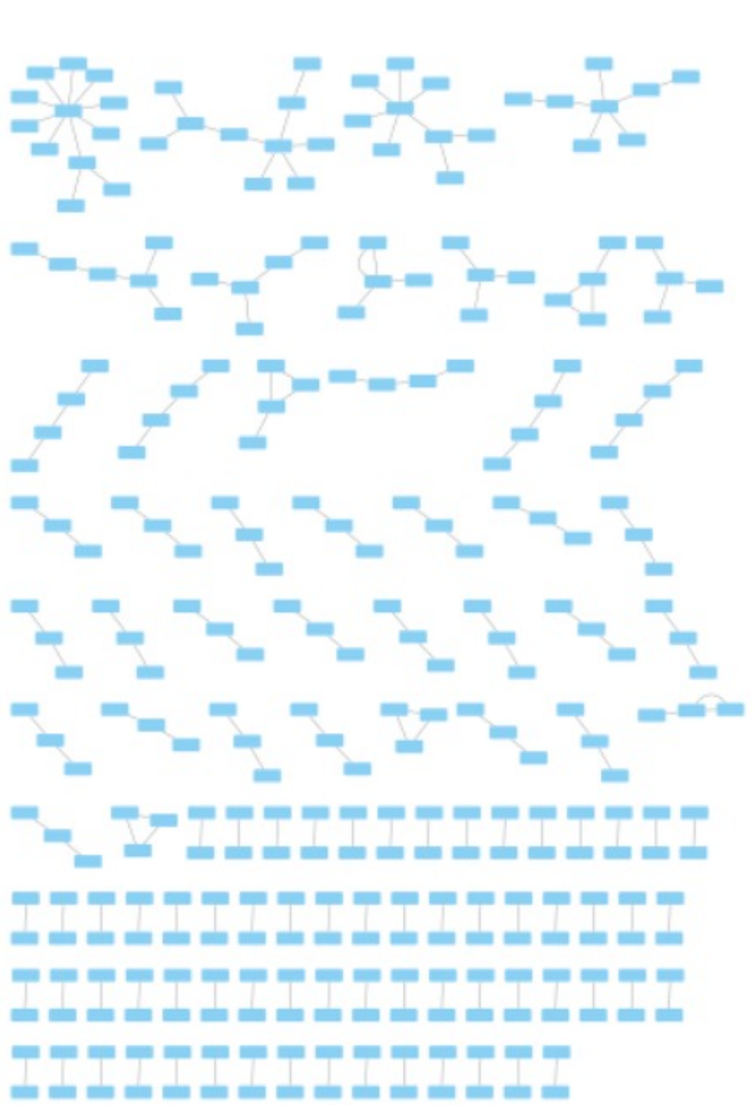
EMBEDDING

While triple extraction is effective at capturing the knowledge explicitly stated in research papers, triple embedding seeks to infer new relations toward completion of the complex and multi-dimensional concepts expressed in research papers. Knowledge graph embeddings are low-dimensional representations of the entities and relations in a knowledge graph. These embeddings provide the context for ideas in a knowledge graph that can be used to infer new relations.

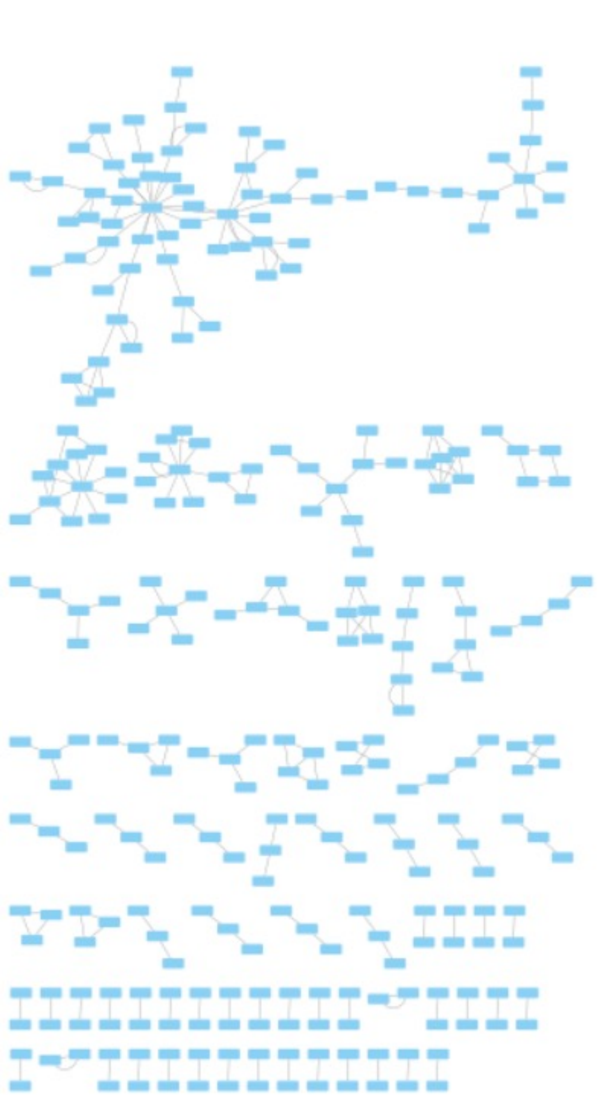
We evaluated several embedding techniques for the task of knowledge graph completion: TransE, TransR, RESCAL, DistMult, and ComplEx. We report the success of these different methods for both the AI-KG and EMA-KN pipelines using mean rank and hit ratio metrics. Results suggest that although RESCAL was the best at the separate tasks of embedding triples from the abstract and full text, TransR was most effective at handling the tasks jointly.

Data	Model	MRR	MR	Hits@1	Hits@3	Hits@10
EMA Abstract	TransE	0.025	**110.0**	0.0	0.021	0.042
	TransR	0.055	130.0	**0.042**	0.042	0.062
	RESCAL	**0.074**	**110.0**	**0.042**	**0.062**	**0.1**
	DistMult	0.022	140.0	0.0	0.021	0.042
	ComplEx	0.032	130.0	0.0	0.042	0.062
EMA Full	TransE	0.0058	**630.0**	0.0	0.0	**0.016**
	TransR	0.0066	650.0	0.0	0.0	0.011
	RESCAL	**0.0098**	710.0	**0.0054**	0.0054	0.011
	DistMult	0.005	640.0	0.0	0.0	0.011
	ComplEx	0.0083	680.0	0.0	**0.011**	**0.016**
EMA all	TransE	0.0025	2100.0	0.0011	0.0011	0.0022
	TransR	**0.019**	2100.0	**0.014**	**0.018**	**0.023**
	RESCAL	0.0035	2300.0	0.0011	0.0022	0.0043
	DistMult	0.0041	2200.0	0.0	0.0022	0.011
	ComplEx	0.01	**2000.0**	0.0054	0.011	0.012
CS abstract	TransE	**0.048**	**120.0**	**0.025**	**0.05**	0.05
	TransR	0.028	140.0	0.0	0.0	**0.075**
	RESCAL	0.041	140.0	**0.025**	0.025	0.05
	DistMult	0.011	130.0	0.0	0.0	0.0
	ComplEx	0.016	130.0	0.0	0.0	0.025
CS full	TransE	0.0017	**1600.0**	0.0	0.0	0.0
	TransR	**0.02**	**1600.0**	**0.016**	**0.019**	**0.023**
	RESCAL	0.008	**1600.0**	0.0018	0.0088	0.012
	DistMult	0.0019	1700.0	0.0	0.0	0.0018
	ComplEx	0.0019	1700.0	0.0	0.0	0.0
CS all	TransE	0.00085	3400.0	0.0	0.0	0.00049
	TransR	**0.024**	**3100.0**	**0.017**	**0.026**	**0.033**
	RESCAL	0.0073	3500.0	0.0054	0.0059	0.0089
	DistMult	0.0012	3300.0	0.0	0.0	0.00099
	ComplEx	0.00097	3400.0	0.0	0.0	0.00049

Computer Science Results



EMA Results



RESULTS EXPLAINED

The previous images show the interconnected nature of the knowledge our system was able to find from datasets of computer science abstracts and EMA related abstracts.

As shown in the previous images, we were able to properly embed relations and entities from both a dataset made from EMA related scientific papers and a dataset made from computer science related papers. The number of edges and nodes found in the computer science dataset was much higher than the number of edges and nodes found in the EMA dataset despite the datasets being the same size, this is most likely due to the fact that our pipeline was initially based on a pipeline that was built specifically for computer science related data. Note how the EMA results do in fact appear to be m**ore connected than the computer science results.**

CONCLUSIONS

Our research suggests that it is indeed possible to create a knowledge graph construction and embedding pipeline free of the need for a dedicated and domain-specific ontology. While the results for the Computer Science baseline were better, the results for the EMA data on the modified pipeline show a lot of promise.

More needs to be done before the EMA-KN pipeline is ready for widespread adoption. We would like to improve our keyword extractor to better prepare the rest of the pipeline for creating complete expressions of the knowledge contained within research papers. We are hopeful these contributions will go a long way on the road toward a Semantic Web.

REFERENCES

- C. Winters, G. Stafford, J. Varghese, F. Zhao, S. Chen, and Y. Shang, “Creation of EMA-KN – A Knowledge Network for Ecological Momentary Assessment,” 2021 IEEE International Conference on Tools with Artificial Intelligence (ICTAI), 2021.
- Dessì D., Osborne F., Reforgiato Recupero D., Buscaldi D., Motta E., Sack H. (2020) AI-KG: An Automatically Generated Knowledge Graph of Artificial Intelligence. In: Pan J.Z. et al. (eds) The Semantic Web – ISWC 2020. ISWC 2020. Lecture Notes in Computer Science, vol 12507. Springer, Cham. https://doi.org/10.1007/978-3-030-62466-8_9

ACKNOWLEDGEMENTS

- This research is supported in part by NSF Grant CNS-1950873 and NIH Grant R01AA027824 and R01AA019546.
- The authors would like to thank the authors of AI-KG for their prior contributions to this field.