Neural Concept Formation in Knowledge Graphs

Agnieszka Dobrowolska
University College London, London

Antonio Vergari
University of California, Los Angeles

Pasquale Minervini
University College London, London

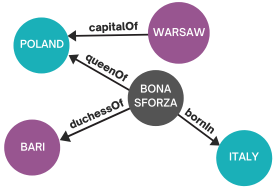
UCL AI Centre

UCLA

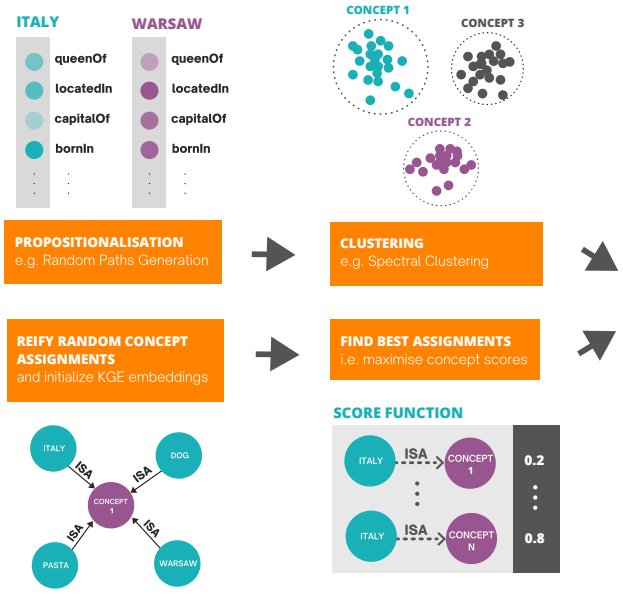
In humans, the ability to summarise knowledge into concepts is believed to play a central role in allowing us to **learn quickly from few examples.**

We propose the task of concept learning in Knowledge Graphs (KG) as a way of **boosting link prediction performance.** Specifically, we propose two algorithms ***ConFormA*** and ***ConFormAE*** for learning concept memberships via **unsupervised clustering** of entities and **explicitly augmenting** the KG with them.

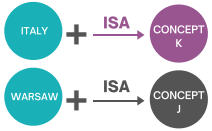
Our approach not only produces **meaningful concepts** but also **improves generalisation for rare predicates.**



▶ Knowledge Graph



▶ E STEP



▶ M STEP

ConFormA and ConFormAE are meta-learners that can leverage ***any KG embedding model out-of-the-box***

Novel concepts and their embedding representations are learned jointly by performing ***Hard EM*** (iterate between E + M steps)

ConFormA

ConFormAE

In Table 1 we compare link prediction performance of a baseline neural link predictor, such as ComplEx or DistMult, with ConFormA and ConFormAE on WN18RR and FB15K-237. In Figure 1 we show the relative improvement upon the baseline for WN18RR, with the largest improvement seen for rare relations. Figure 2 shows fragments of propotytpical concept clusters formed for UMLS using ConFormAE.

Table 1: Mean reciprocal rank (MRR) and Hits (H) at 1,3,10 for ConFormA and ConFormAE when using DistMult or ComplEx as baseline KGE models on WN18RR and FB15K-237 KGs for different values of embedding sizes (k). Best values for each metric and k in bold.

	k	Model	DistMult MRR	DistMult H@1	DistMult H@3	DistMult H@10	ComplEx MRR	ComplEx H@1	ComplEx H@3	ComplEx H@10
WN18RR	500	Baseline	44.40	40.30	45.63	52.81	47.85	43.41	49.44	56.57
		ConFormA	44.19	39.96	45.33	52.94	48.55	43.94	50.24	**57.67**
		ConFormAE	**44.89**	**40.84**	**46.00**	**53.38**	**48.77**	**44.42**	**50.48**	57.29
	1000	Baseline	44.80	41.02	45.80	52.52	48.34	43.81	50.18	56.99
		ConFormA	**45.47**	**41.16**	**46.76**	**54.28**	49.12	44.42	**50.83**	**58.70**
		ConFormAE	45.20	40.97	46.41	54.05	**49.25**	**44.81**	50.81	58.33
	2000	Baseline	45.20	41.05	**46.39**	53.75	48.62	44.07	50.34	57.28
		ConFormA	44.93	40.60	45.98	53.67	**49.40**	44.80	**50.86**	**59.00**
		ConFormAE	**45.38**	**41.16**	**46.39**	**54.04**	49.42	**45.20**	50.41	58.42
FB15K237	500	Baseline	34.88	25.56	38.34	53.52	35.89	26.47	39.31	54.82
		ConFormA	34.92	25.65	38.39	53.54	36.08	**26.77**	39.39	55.00
		ConFormAE	**35.01**	**25.72**	**38.40**	**53.65**	**36.13**	26.76	**39.46**	**55.09**
	1000	Baseline	35.26	25.83	**38.82**	54.26	36.18	26.69	**39.81**	55.21
		ConFormA	35.30	25.91	38.78	**54.28**	36.26	**26.88**	39.74	55.22
		ConFormAE	**35.40**	**26.11**	38.77	**54.28**	**36.27**	26.84	**39.81**	**55.34**
	2000	Baseline	35.47	26.13	38.76	54.42	**36.37**	**27.01**	**39.89**	**55.45**
		ConFormA	35.55	26.18	**39.03**	54.32	**36.37**	26.99	**39.89**	55.19
		ConFormAE	**35.62**	**26.31**	39.02	**54.43**	36.34	26.96	39.81	55.36

Figure 1

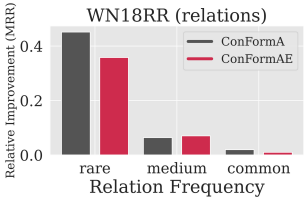


Figure 2

Concept 1	Concept 2	Concept 3	Concept 4	Concept 5
• alga • amphibian • animal • virus • bacterium • bird • invertebrate	• cell_or_molecular_dysfunction • disease_or_syndrome • experimental_model_of_disease • injury_or_poisoning • mental_or_behavioral_dysfunction • neoplastic_process • pathologic_function.	• antibiotic • vitamin • substance • enzyme • food • hormone • organic_chemical	• activity • behavior • health_care_activity • daily_or_recreational_activity • educational_activity • individual_behavior • laboratory_procedure	• biologic_function • cell_function • genetic_function • mental_process • molecular_function • organ_or_tissue_function • natural_phenomenon_or_process