NeuSymEA: Neuro-symbolic Entity Alignment via Variational Inference

Shengyuan Chen, Zheng Yuan, Qinggang Zhang, Wen Hua, Jiannong Cao, Xiao Huang

Background

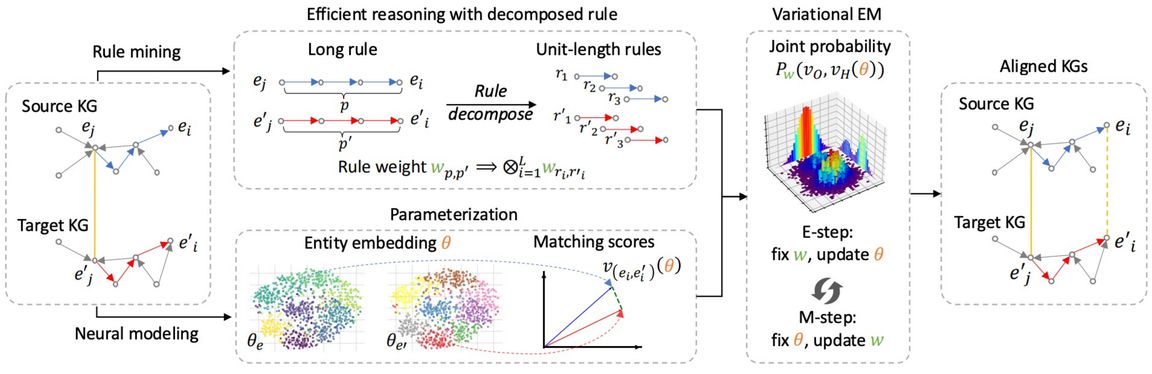
Entity alignment (EA) Task: to merge two knowledge graphs by identifying equivalent entity pairs.

Existing work: can be categorized into symbolic and neural models. Symbolic models, while precise, struggle with substructure heterogeneity and sparsity, whereas neural models, although effective, generally lack interpretability and cannot handle uncertainty.

Framework – NeuSymEA

NeuSymEA models the joint probability of all possible pairs' truth scores in a Markov random field, regulated by a set of rules, and optimizes it with the variational EM algorithm

- **E-step**: a neural model parameterizes the truth score distributions and infers missing alignments.
- **M-step**: the rule weights are updated based on the observed and inferred alignments, handling uncertainty
- **Extend to long rules**: We introduce an efficient symbolic inference engine driven by logic deduction, enabling reasoning with long rules.

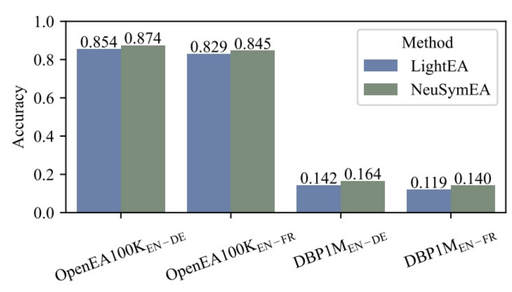


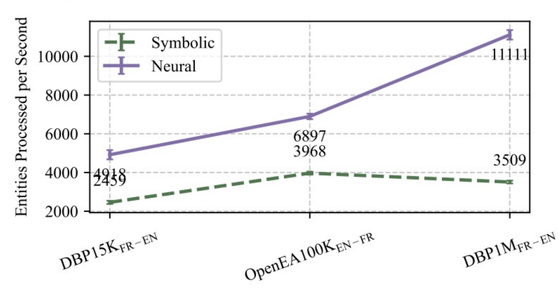
Framework illustration of NeuSymEA.

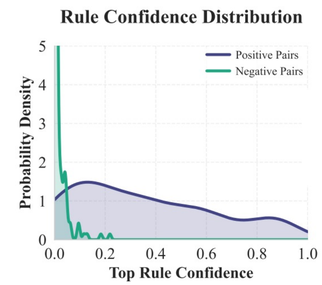
Experimental Results

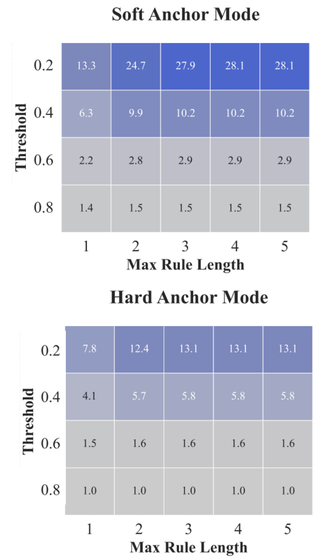
Category	Model	JA-EN Hit@1	JA-EN Hit@10	JA-EN MRR	FR-EN Hit@1	FR-EN Hit@10	FR-EN MRR	ZH-EN Hit@1	ZH-EN Hit@10	ZH-EN MRR
Results on the full DBP15K dataset										
Neural	GCNAlign	0.221	0.461	0.302	0.205	0.475	0.295	0.189	0.438	0.271
	BootEA	0.454	0.782	0.564	0.443	0.799	0.564	0.486	0.814	0.600
	AlignE	0.356	0.715	0.476	0.346	0.731	0.475	0.333	0.690	0.453
	Dual-AMN	0.627	0.883	0.717	0.652	0.908	0.744	0.650	0.884	0.732
	LightEA	0.736	0.894	0.793	0.782	0.919	0.832	0.725	0.874	0.779
Symbolic	PARIS	0.589	-	-	0.618	-	-	0.603	-	-
Neuro-symbolic	PRASE	0.611	-	-	0.647	-	-	0.652	-	-
Ours	NeuSymEA-D	**0.806**	**0.942**	**0.855**	0.827	**0.952**	**0.871**	**0.801**	**0.925**	**0.843**
	NeuSymEA-L	0.781	0.907	0.826	**0.834**	0.937	**0.871**	0.785	0.894	0.825

Scalability on large scale KGs. (Left) Hit@1 alignment performance on large KGs. (Right) Per-second processed entities of neural and symbolic components on different scales of KGs.









Robustness in low-resource settings, achieving 73.7% hit@1 accuracy on DBP15KFR-EN with only 1% pairs as seed alignments.

Dataset	Model	1% H@1	1% H@10	1% MRR	5% H@1	5% H@10	5% MRR	10% H@1	10% H@10	10% MRR	20% H@1	20% H@10	20% MRR
JA-EN	AlignE	0.007	0.034	0.016	0.080	0.268	0.143	0.244	0.588	0.356	0.433	0.783	0.552
	BootEA	0.010	0.040	0.021	0.379	0.683	0.481	0.468	0.779	0.573	0.530	0.829	0.631
	GCNAlign	0.029	0.128	0.063	0.127	0.368	0.206	0.209	0.515	0.310	0.331	0.662	0.443
	PARIS	0.145	-	-	0.340	-	-	0.450	-	-	0.565	-	-
	PRASE	0.163	-	-	0.432	-	-	0.508	-	-	0.580	-	-
	Dual-AMN	0.239	0.519	0.334	0.509	0.795	0.611	0.652	0.887	0.738	0.750	0.927	0.815
	RREA	0.253	0.486	0.332	0.558	0.830	0.653	0.672	**0.903**	0.756	0.789	**0.956**	0.853
	LightEA	0.291	0.514	0.363	0.627	0.806	0.689	0.714	0.874	0.771	0.778	0.911	0.828
	EMEA	0.411	-	0.488	0.630	-	0.710	0.688	-	0.764	0.736	-	0.807
	PEEA	0.242	0.519	0.333	0.490	0.785	0.589	0.612	0.834	0.679	0.703	0.912	0.777
	NeuSymEA-D	0.481	0.684	0.550	0.692	0.855	0.749	0.742	0.895	0.796	0.835	0.953	0.879
	NeuSymEA-L	**0.632**	**0.779**	**0.683**	**0.733**	**0.870**	**0.781**	**0.773**	0.900	**0.818**	**0.858**	0.954	**0.894**