SAFRAN: An interpretable, rule-based link prediction method outperforming embedding models

Simon Ott[1], Christian Meilicke[2], Matthias Samwald[1]

[1] Institute of Artificial Intelligence and Decision Support, Medical University of Vienna, Austria
[2] Data and Web Science Research Group, University Mannheim, Germany

AUTOMATED KNOWLEDGE BASE CONSTRUCTION

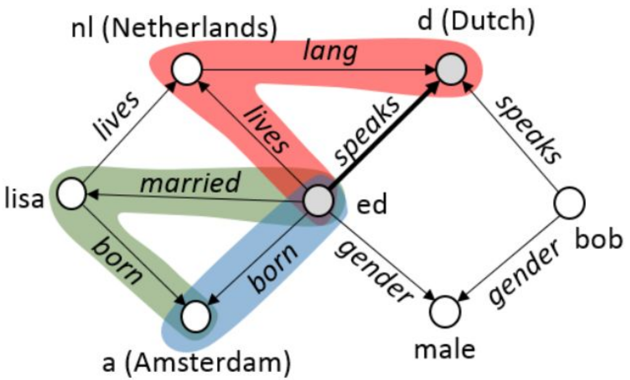
Rule-based approach AnyBURL

Advantages

- Efficient rule learning
- Explainable!

Limitations

- Functional redundant rules
- Slow rule application

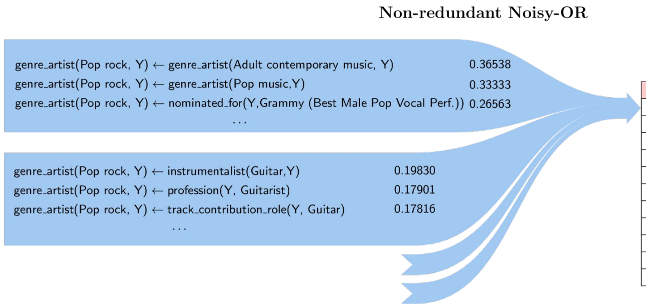


Meilicke, C. et al. (2019). Anytime Bottom-Up Rule Learning for Knowledge Graph Completion

$$speaks(X, Y) \leftarrow lives(X, A_2), lang(A_2, Y)$$
$$speaks(X, d) \leftarrow married(X, A_2), born(A_2, a)$$

SAFRAN

- Scalable and fast non-redundant rule application
- **Clustering of redundant rules**
  Elevating predictive performance
- **Scalability**
  Enables the application of symbolic rules to large scale KBs



genre_artist(?, Bryan Adams)

Entity	Confidence	
Pop rock	0.99817	✓
Country music	0.95141	
Blues rock	0.94286	
Dance-pop	0.93301	
Rythm and blues	0.93282	
Psychedelic rock	0.91512	
Folk rock	0.88181	
Contemporary R&B	0.86727	
Rock and roll	0.86479	
Progressive rock	0.84606	
Jazz	0.83302	

https://github.com/OpenBioLink/SAFRAN

Results

	Approach			FB15K-237 MRR	FB15K-237 hits@1	FB15K-237 hits@10	WN18RR MRR	WN18RR hits@1	WN18RR hits@10		YAGO3-10 MRR	YAGO3-10 hits@1	YAGO3-10 hits@10
Latent	RESCAL		⋆	.357	.263	.541	.468	.439	.521				
	TransE		⋆	.313	.221	.497	.227	.053	.526	¶	.501	.405	.673
	DistMult		⋆	.343	.250	.531	.452	.413	.531	¶	.501	.412	.661
	ComplEx		⋆	.348	.253	.534	.477	.438	.543	¶	**.576**	**.505**	**.704**
	ConvE		⋆	.339	.248	.521	.447	.411	.508	¶	.488	.399	.657
	RotatE		¶	.336	.238	.531	.475	.426	.574	¶	.498	.405	.670
	TuckER		¶	.352	.259	.536	.459	.430	.514	¶	.544	.465	.680
	HAKE		▽	.346	.250	**.542**	.497	.452	**.582**		.545	.462	.694
Interpretable	C-NN		△	.296	.222	.446	.469	.444	.519				
	DRUM		‡	*.343*†	*.255*†	*.516*†	*.486*†	*.425*†	*.586*†				
	Neural LP		♢	*.240*†		*.362*†	*.435*†	*.371*†	*.566*†				
	GPFL		♢	.322	.247	.504	.480	.449	.552				
	AMIE+		♣		.174	.409		.358	.388				
	RuleN		♣		.182	.420		.427	.536				
	RLvLR		♯	.240		.393							
	AnyBURL with SAFRAN	Maximum		.355	.270	.519	.494	.452	.572		.559	.487	.686
		Noisy-OR		.342	.258	.502	.454	.399	.562		.524	.444	.672
		VS		.363	.277	.526	.497	.455	.572		.560	.489	.689
		NRNO * (grid)		.370	.287	.531	.501	.457	.581		.564	.492	.691
		NRNO * (random)		**.389**	**.298**	.537	**.502**	**.459**	.578		.564	.491	.693

This work received funding from netidee grant 5171 (’OpenBioLink’).