LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

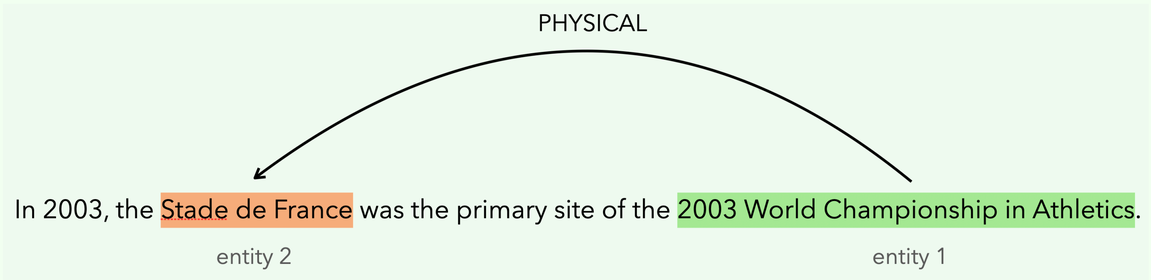
IT UNIVERSITY OF COPENHAGEN

Danish Data Science Academy

How to Encode Domain Information in Relation Classification

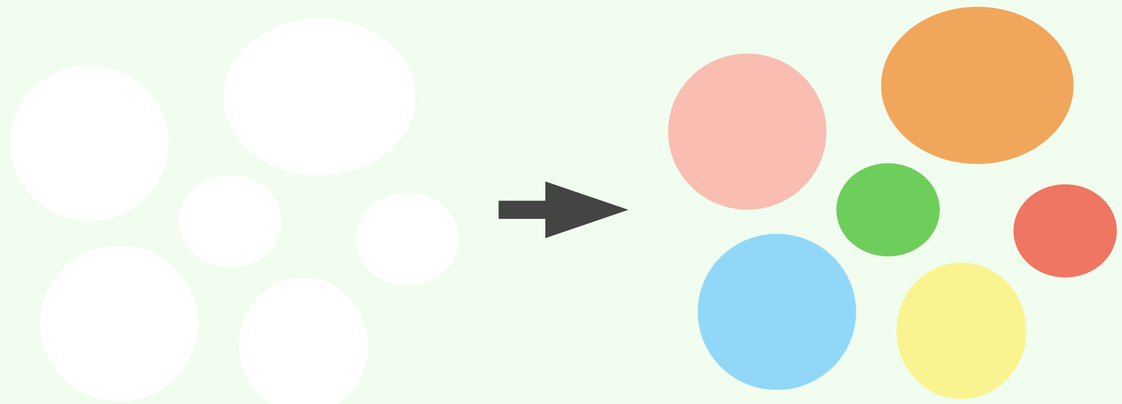
Elisa Bassignana, Viggo Unmack Gascou, Frida Nøhr Laustsen, Gustav Kristensen, Marie Haahr Petersen, Rob van der Goot, Barbara Plank

THE TASK OF RELATION CLASSIFICATION



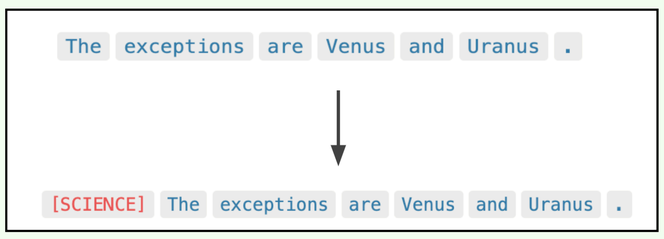
THE PROBLEM

How to maximize the training data without losing domain-specific information?

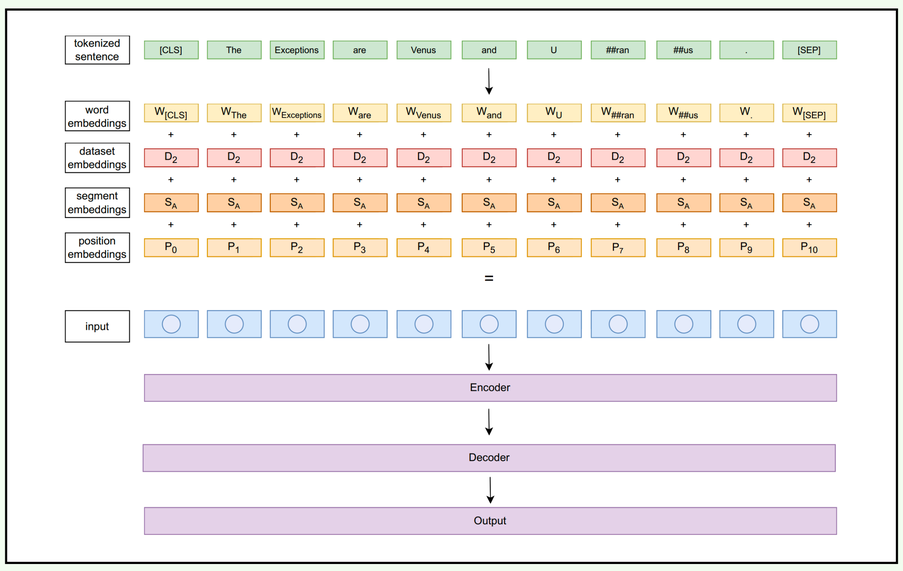


MODEL VARIATIONS

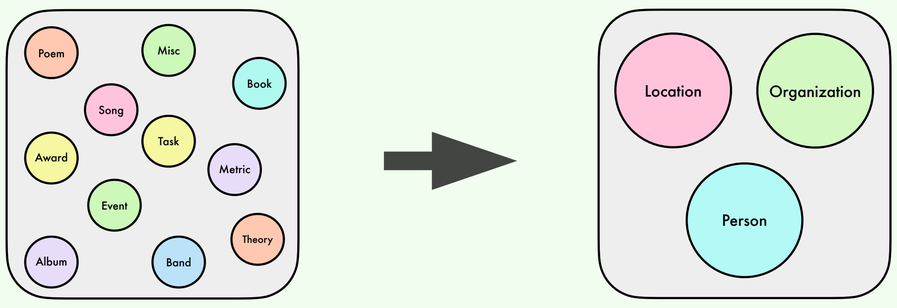
Special Domain Markers



Dataset Embeddings



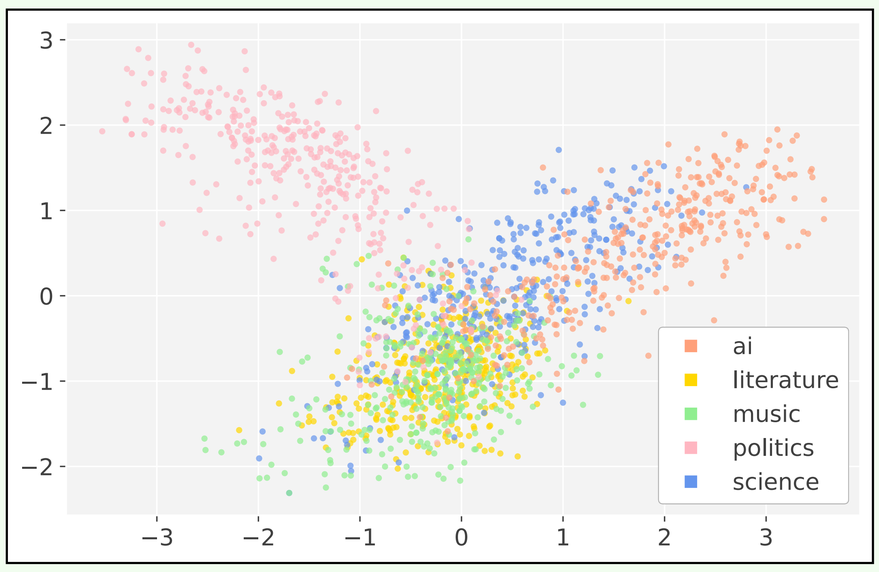
Domain Specific Entity Types



<E1:person> Steve Jobs </E1:person> was born in <E2:location> San Francisco </E2:location>

ANALYSIS OF DOMAINS

PCA plot of the untrained embeddings of the instances in the development set, colored by domain



RESULTS

								avg.
dev	Baseline	25.45	31.35	39.46	39.69	38.84	38.09	35.48
	Dataset Emb.	15.38	22.22	24.77	32.64	30.95	29.80	25.96
	Domain Mark.	26.36	32.77	40.31	42.65	40.59	38.71	**36.90**
	Fine-grain.	23.67	32.67	35.35	38.76	38.23	35.94	34.10
	Coarse-grain.	24.46	31.56	38.59	39.33	38.09	37.90	34.99
	Dom. + Coarse	24.52	32.02	39.63	42.19	40.01	37.17	35.92
test	Baseline	24.73	34.12	39.67	39.96	44.64	35.71	36.47
	Domain Mark.	26.72	37.62	43.57	41.48	44.88	37.69	**38.66**

Macro-F1 scores of the explored setups. Dom. + Coarse indicates the combination of special domain markers with the coarse-grained entity types.

ANALYSIS OF RELATION TYPES

PCA plot of the trained embeddings of the most frequent relation labels in the development set, colored by relation labels and shaped by domain