

Stranger than paradigms

Word Embedding Benchmarks Don't Align With Morphology



THE OHIO STATE UNIVERSITY

Timothee Mickus, University of Helsinki Maria Copot, The Ohio State University

What we do

RQ: We use word embeddings for morphology studies, but do we evaluate and train them properly?

- We design benchmarks to evaluate word2vec SGNS embeddings for French morphology
- We conduct an extensive hyperparameter exploration across 1200 models
- We study how hyperparameter choices affect performance on morphology as well as traditional NLP benchmark

Take-home messages

- The negative sampling exponent is a **crucial** hyperparameter: it must be **high** for inflection but **low** for NLP lexical semantic tasks.

A high negative sampling exponent forces us to distinguish all embeddings from frequent (and morphologically ambiguous) words. A low negative sampling exponent forces us to distinguish all embeddings from infrequent (and semantically varied) words.

- The optimal SGNS embeddings for morphology are sub-optimal for commonplace NLP benchmarks & vice-versa.

Morphology or semantics?

Performance on morphology tasks anticorrelates with performance on NLP lexical semantics benchmarks

	PCC	SCC	PCP	SCP	DerPS	DerPF	POS	DerCF	GATS/D	GATS/I	GATS/sem	FEEL	simlex	DerCS
PCC	1	0.98	0.77	0.7	0.48	0.37	0.43	0.21	-0.59	-0.57	-0.73	-0.43	-0.69	-0.49
SCC	0.98	1	0.73	0.62	0.35	0.23	0.3	0.14	-0.67	-0.65	-0.79	-0.52	-0.72	-0.59
PCP	0.77	0.73	1	0.71	0.53	0.47	0.46	0.18	-0.26	-0.22	-0.43	-0.17	-0.52	-0.29
SCP	0.7	0.62	0.71	1	0.7	0.65	0.66	0.37	-0.21	-0.12	-0.33	-0.04	-0.39	-0.11
DerPS	0.48	0.35	0.53	0.7	1	0.97	0.96	0.48	-0.06	-0.02	0.2	0.12	-0.38	0.2
DerPF	0.37	0.23	0.47	0.65	0.97	1	0.95	0.41	0.06	0.1	-0.07	0.2	-0.31	0.29
POS	0.43	0.3	0.46	0.66	0.96	0.95	1	0.4	-0.1	-0.05	-0.21	0.06	-0.42	0.2
DerCF	0.21	0.14	0.18	0.37	0.48	0.41	0.4	1	0.14	0.16	0	0.37	0.25	0.49
GATS/D	-0.59	-0.67	-0.26	-0.21	-0.06	0.06	-0.1	0.14	1	0.94	0.94	0.87	0.79	0.75
GATS/I	-0.57	-0.65	-0.22	-0.12	-0.02	0.1	-0.05	0.16	0.94	1	0.94	0.84	0.78	0.78
GATS/sem	-0.73	-0.79	-0.43	-0.33	-0.2	-0.07	-0.21	0	0.94	0.94	1	0.8	0.83	0.72
FEEL	-0.43	-0.52	-0.17	-0.04	0.12	0.2	0.06	0.37	0.87	0.84	0.8	1	0.73	0.76
simlex	-0.69	-0.72	-0.52	-0.39	-0.38	-0.31	-0.42	0.25	0.79	0.78	0.83	0.73	1	0.67
DerCS	-0.49	-0.59	-0.29	-0.11	0.2	0.29	0.2	0.49	0.75	0.78	0.72	0.76	0.67	1

