HyperEmbed

Tradeoffs Between Resources and Performance in NLP Tasks with Hyperdimensional Computing enabled embedding of **n**-gram statistics

•••

Pedro Alonso*, **Kumar Shridhar***, Denis Kleyko*, Evgeny Osipov, Marcus Liwicki

PML4DC@ICLR 2020

April 2020

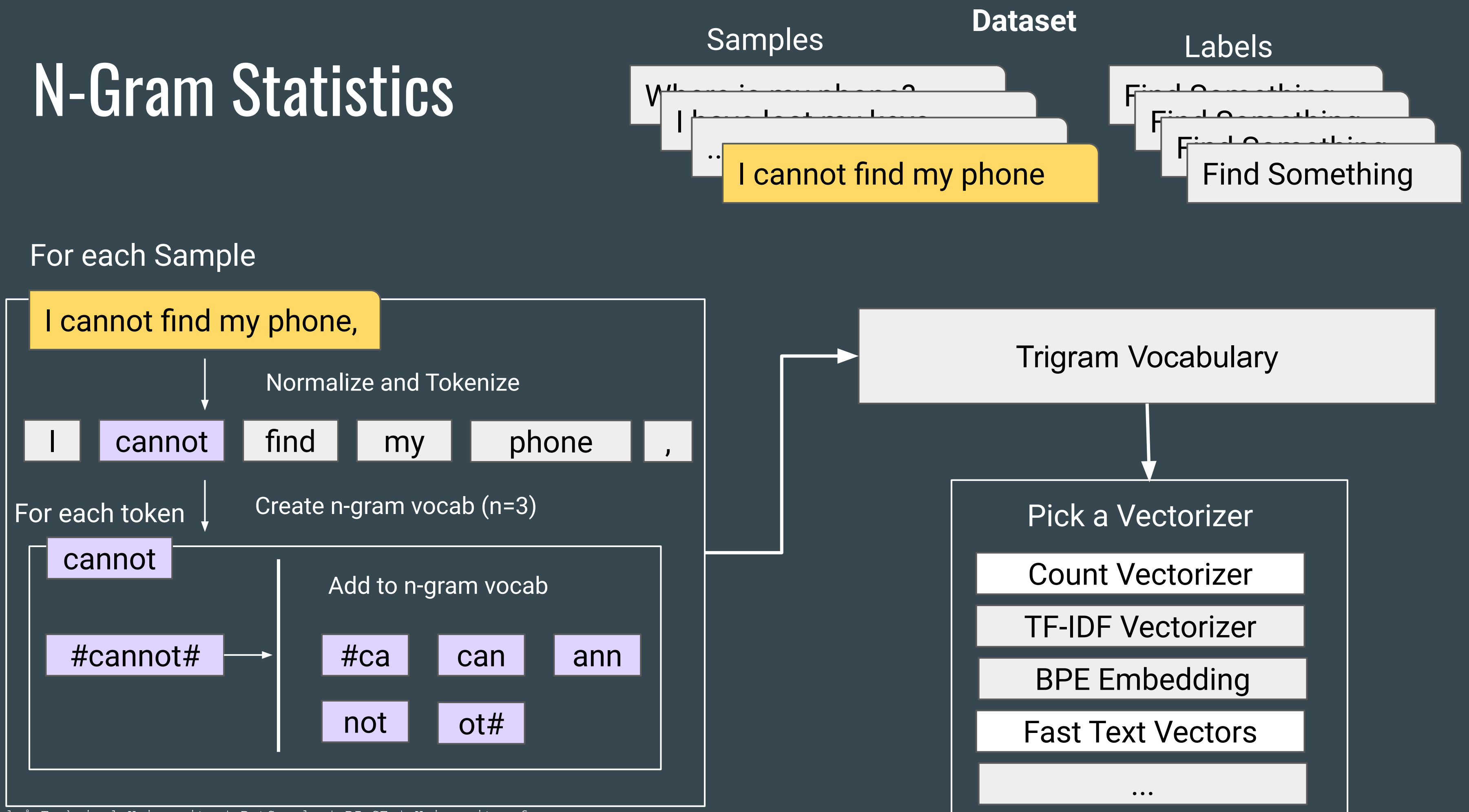
Luleå Technical University | BotSupply | RI.SE | University of Berkeley

Main Idea

Recent advances in Deep Learning have led to a significant performance increase on several NLP tasks, however, the models become more and more computationally demanding.

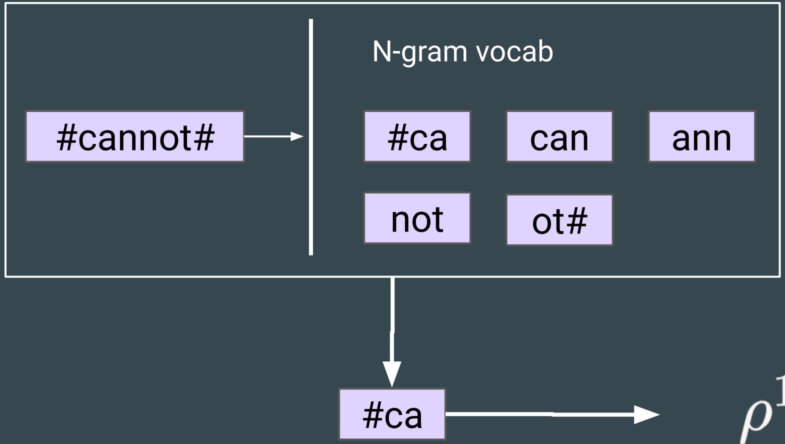
This **paper** tackles the **domain of computationally efficient algorithms** for **NLP** tasks.

N-Gram Statistics



Embedding into a HD Vector

1. Every symbol is assigned a ***d***-dimensional HD vector.
2. These vectors are stored in matrix **H**.
3. A unique HD vector for each N-gram is formed using binding and permutation operations
4. All observed N-gram are stored in a single HD vector using the bundling operation



$$\rho^1(H_\#) \odot \rho^2(H_c) \odot \rho^3(H_a)$$

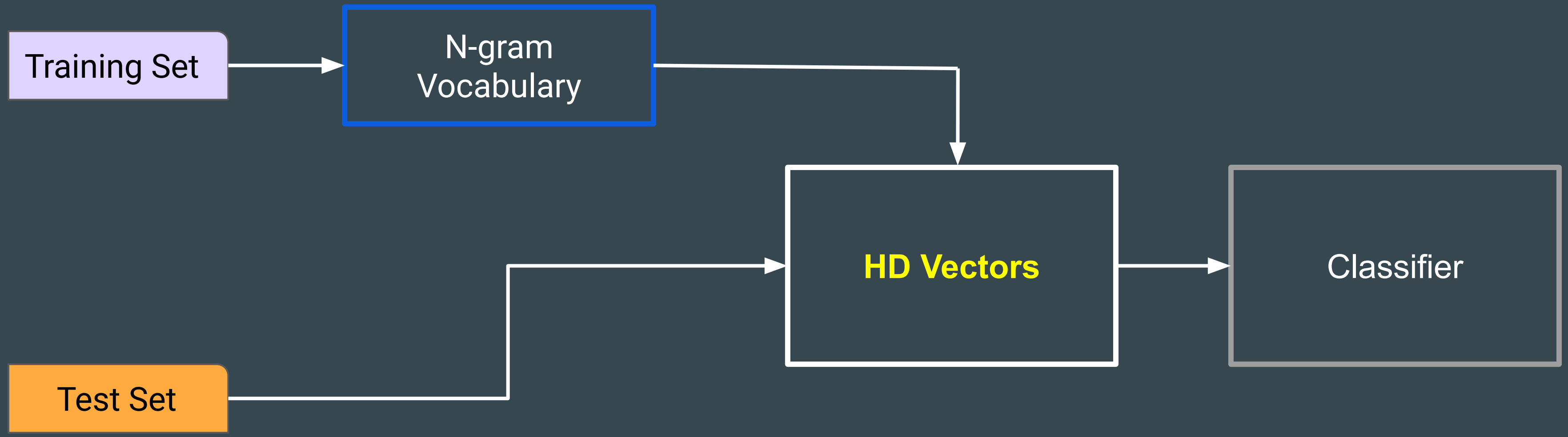
Three Key operation

01 **Bundling** + (position wise addition)

02 **Binding** $\odot$ (position wise multiplication)

03 **Permutation** ρ

Full NLP Classification pipeline



Dataset

Small Dataset

- Ask Ubuntu : **2** Intents (50 train, 50 test samples per intent)
- Chatbot Corpus : **5** Intents (10 train, 20 test samples per intent)
- WebApplication Corpus : **8** intents (3 train, 7 test samples per intent)

Large Dataset

- 20 News Corpus : **20** Intents (11,500 train, 7,500 test samples per intent)

Empirical Evaluation: ML algorithms

AskUbuntu Corpus. *N*=512

	F_1 score			Resources: SH vs. HD			Resources: SH vs. BPE		
Classifier	SH	BPE	HD	Tr.	Ts.	Mem.	Tr.	Ts.	Mem.
MLP	0.92	0.91	0.91	4.62	3.84	6.18	1.67	1.61	1.72
Passive Aggr.	0.92	0.93	0.90	4.86	3.07	6.31	2.19	2.14	1.76
SGD Classifier	0.89	0.89	0.88	4.66	3.50	6.31	1.94	2.16	1.76
Ridge Classifier	0.90	0.91	0.90	3.91	4.74	6.31	1.63	1.62	1.76
KNN Classifier	0.79	0.72	0.82	2.11	4.53	8.48	1.56	1.79	1.76
Nearest Centroid	0.90	0.89	0.90	1.66	3.41	6.32	1.35	1.87	1.76
Linear SVC	0.90	0.92	0.90	1.18	2.39	6.29	0.91	1.91	1.76
Random Forest	0.88	0.90	0.86	0.91	1.09	6.11	1.15	0.96	1.75
Bernoulli NB	0.91	0.92	0.85	2.30	3.72	6.34	1.96	2.42	1.76

On par F_1 score with **huge reduction** in **space** and **time** complexity.

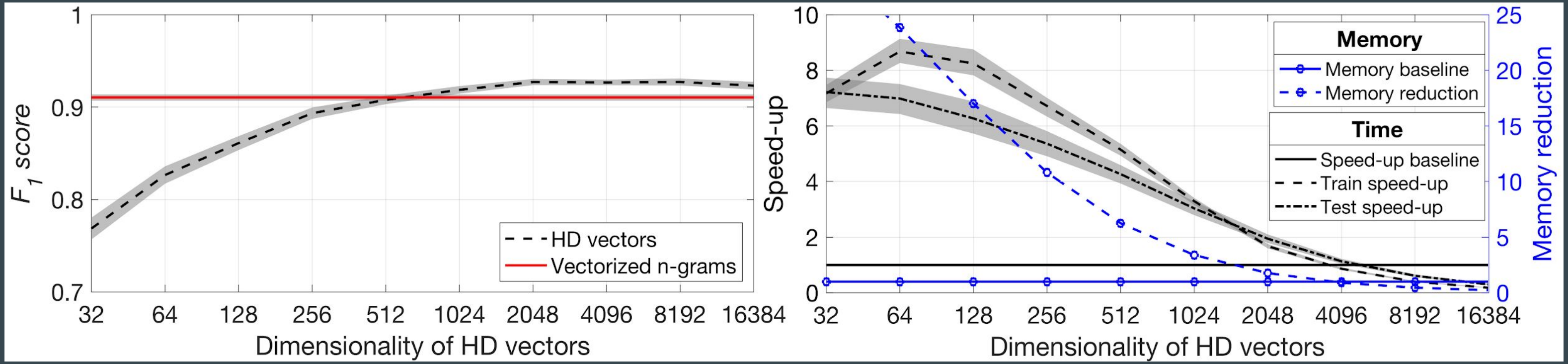
Empirical Evaluation: ML algorithms

20 News Corpus. *N*=512

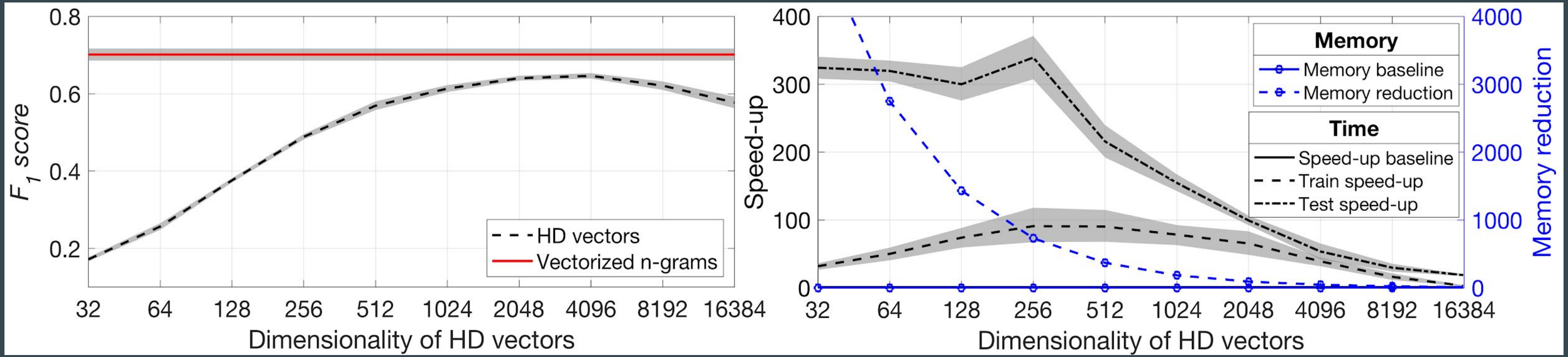
	F_1 score		Resources: SH vs. HD		
Classifier	SH	HD	Train speed-up	Test speed-up	Memory reduction
MLP	0.72	0.64	53.23	79.50	93.19
Passive Aggr.	0.74	0.69	103.64	202.95	93.42
SGD Classifier	0.70	0.66	105.43	186.31	93.42
Ridge Classifier	0.16	0.71	45.46	338.01	93.42
KNN Classifier	0.31	0.31	184.47	65.87	127.54
Nearest Centroid	0.08	0.15	212.75	254.74	93.42
Linear SVC	0.75	0.69	5.11	176.62	93.42
Random Forest	0.58	0.26	4.27	21.43	93.41
Bernoulli NB	0.60	0.15	57.72	56.54	93.42

On par F_1 score with **huge reduction** in **space** and **time** complexity.

Empirical Evaluation: HyperEmbed vs. CountVectorizer (MLP)



Chatbot Corpus



20 News Corpus

Thank You!

Paper: https://arxiv.org/abs/2003.01821