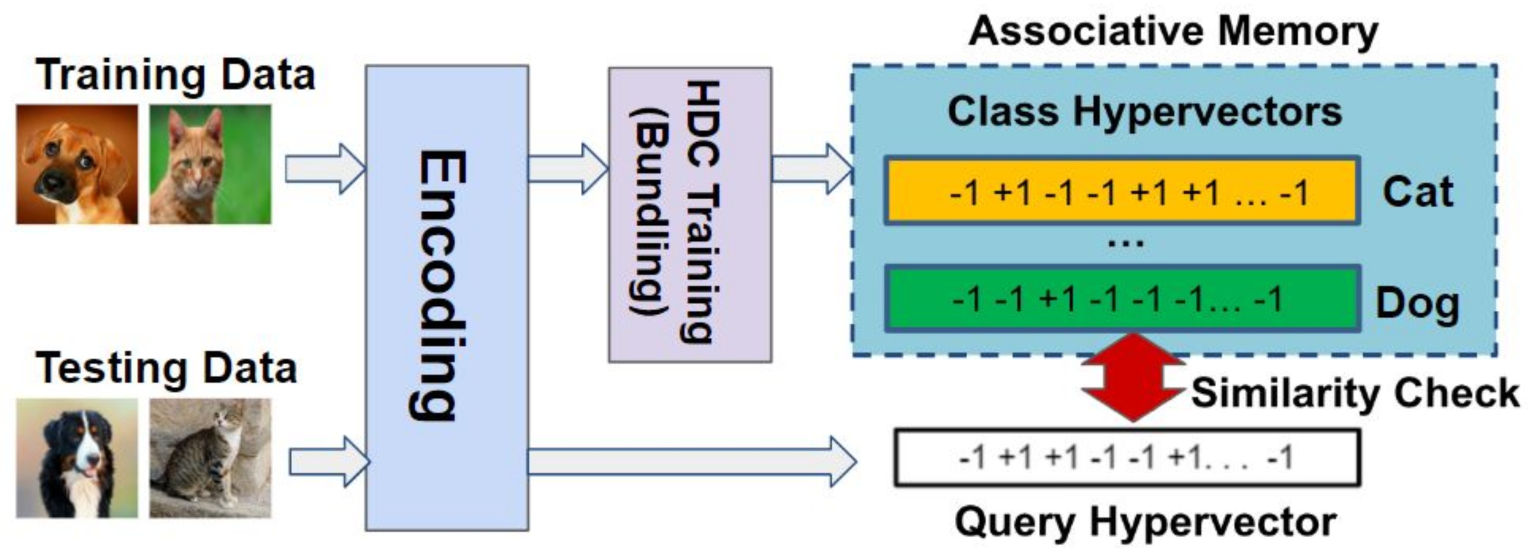
Unleashing Hyperdimensional Computing with Nyström Method based Encoding

Quanling Zhao, Anthony Thomas, Ari Brin, Xiaofan Yu, Tajana Rosing
Department of Computer Science and Engineering, University of California San Diego

Background and Motivations



Success of HDC based machine learning approaches is heavily dependent on the encoding function that maps raw data to high-dimensional space

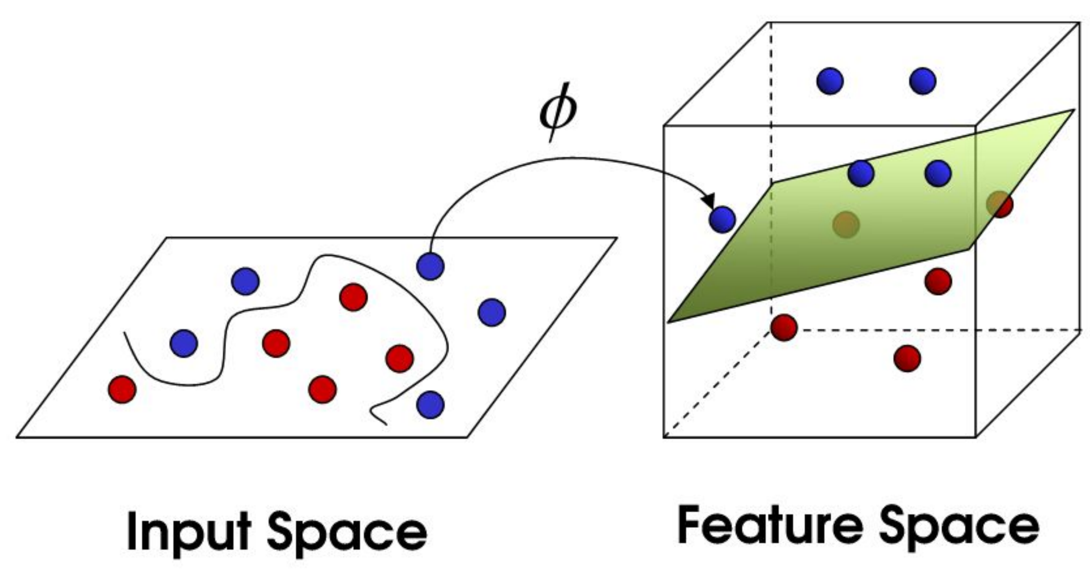
Hyperdimensional Computing (HDC)

- Lightweight and efficient computing paradigm capable of ML tasks such as classification, regression.
- Amenable to highly-parallel circuitry, and require low-precision processing.

Limitations of Existing HDC encoding methods

- Tend to only capture basic notion of similarities, which may be suboptimal when dealing with data with complex structure (e.g. strings, graphs)
- Existing kernel method literature has wide variety of kernel functions (similarity functions)

Kernel methods and HDC



- Kernel method solves "nonlinear" task using linear model with the help of kernel functions.
- Just like in HDC, kernel methods work by embedding data into a high-dimensional space wherein similarities are measured using inner-products

- **Inner-products in HD space should be reflective of some salient notion of similarity on ambient space.**
- **Idea: Construct HD encoding functions using suitable kernel functions.**

Random Fourier Features (RFF)

- Commonly referred as "non-linear" encoding in HDC literature.
- Capable of modeling shift-invariant kernels in HDC (e.g. the Gaussian kernel, polynomial kernel)

Why Nyström method ?

$$\hat{G_{ij}} \approx \phi_{nys}(x_i)^T \phi_{nys}(x_j) = \left(\Lambda^{-\frac{1}{2}} Q^T C^{(i)}\right)^T \left(\Lambda^{-\frac{1}{2}} Q^T C^{(j)}\right)$$

- RFF only works with shift-invariant kernels on a Euclidean space, which many useful kernels do not satisfy (i.e. kernels on graphs and strings)

Nyström Method based HDC Encoding & Main Results

- **"Top down"** approach where embedding directly approximates some data-appropriate notion of similarity.

Theorem 1 *Define $\Theta_i = \Lambda^{-\frac{1}{2}} Q^T C^{(i)}$ and $\Theta_i \in \mathbb{R}^n$. Based on above HDC encoding algorithm, the encoding function $\phi : \mathcal{X} \to \mathcal{H}$ can be write as following:*

$$\phi(x_i) = \sqrt{\frac{\pi}{2d}} sign(P_{rp}\Theta_i) \tag{3}$$

We pose no restriction on Kernel K, the following holds up to a first order approximation:

$$\mathbb{E}\left[\langle \phi(x_i), \phi(x_j) \rangle\right] \approx \frac{\hat{G}_{ij}}{\sqrt{\hat{G}_{ii}\hat{G}_{jj}}} \quad \textit{i.e. normalized kernel} \tag{4}$$

Where $\hat{G}_{ij}$ is estimated kernel value between x_i and x_j produced by Nyström method and the expectation is taken with respect to randomness and orthogonality in P_{rp}.

- our proposed encoding method preserves the kernel in HD space of some user defined kernel..

Experimental setup:

Task	dataset	# of training samples	# of testing samples	# of classes
String	Protein sequences [39]	721	181	6
	SMS Spam collection [1]	4459	1115	2
Graph	ENZYME [1] [3]	480	120	6
	NCI1 [2] [45]	3288	822	2
Image	MNIST [23]	60000	10000	10
	FashionMNIST [47]	60000	10000	10

Key Results:

Task	Dataset	Ours	N-gram [13]	GraphHD [31]	LinearRP [34, 12]	NonlinearRP [44]
String	Protein sequence	**99%**	81%	-	-	-
	SMS Spam collection	**96%**	93%	-	-	-
Graph	ENZYME	**63%**	-	26%	-	-
	NCI1	**72%**	-	62%	-	-
Image	MNIST	96%	-	-	96%	**97%**
	FashionMNIST	**86%**	-	-	85%	**86%**

Summary & Acknowledgement

- In summary, we propose a new way to generate embeddings for HDC which can turn any user-defined positive-semidefinite similarity function into an equivalent embedding.
- This work allows future HDC works to exploit the power of kernel methods while still conforming to the general formalism and benefits of HDC.

Discussion & Future Work

- We recognize that the improvements in our proposed HDC encoding methods also come with additional computation costs in the form of kernel evaluation.
- How to achieve the best efficiency-accuracy trade-offs for HDC applications are non-trivial problems that need further investigations.

Acknowledgement: This work was supported in part by National Science Foundation under Grants, #1826967, #2100237, #2112167, #1911095, #2112665, and in part by SRC under task #3021.001. This work was also supported in part by PRISM and CoCoSys, centers in JUMP 2.0, an SRC program sponsored by DARPA.