Towards a topological-geometrical theory of GENOEs for data analysis and machine learning

刘健

拓扑与几何技术创新中心
河北师范大学

September 22, 2021

目录

Retrospect: Mathematical Model

- Data: $\Phi = \mathrm{map}(X, \mathbb{R})$, where X is a topological space.
- Action: a topological group G action on Φ (with respect to the topology induced by the pseudo-metric given by betti numbers).
- GENESs: group equivariant non-expansive operators such as convolution (two reasons).
- Features: A familay of GENESs are learned as features.

Machine Learning

- Model: GENEOs based on topological metric in TDA.
- Prior knowledge: dateset with G-invariant pseudo-metric.

$$\hat{d}(\phi_1, \phi_2) = \hat{d}(\phi_1 \circ g, \phi_2) = \hat{d}(\phi_1, \phi_2 \circ g) = \hat{d}(\phi_1 \circ g, \phi_2 \circ g).$$

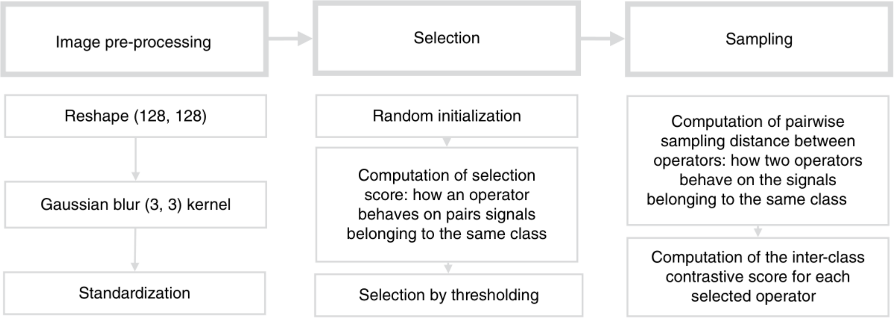
- Features: a family of mininal, contrastive and efficient GENEOs $\mathcal{F}$.
- Classification: for a sample $\phi \in \Phi$, consider $F(\phi)$ for all $F \in \mathcal{F}$.

Machine Learning

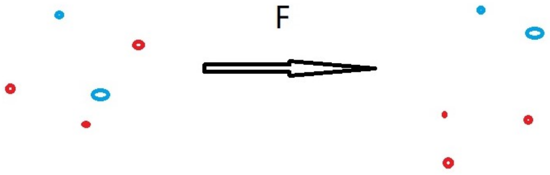
The idea is essentially a generalization of CNN.

- Convolution operators $\rightarrow$ GENEOs
- Group action of Euclidean plane isometries $\rightarrow$ Group action of G-invariant pseudo-metric
- Convolution kernels $\rightarrow$ a family of suitable GENEOs

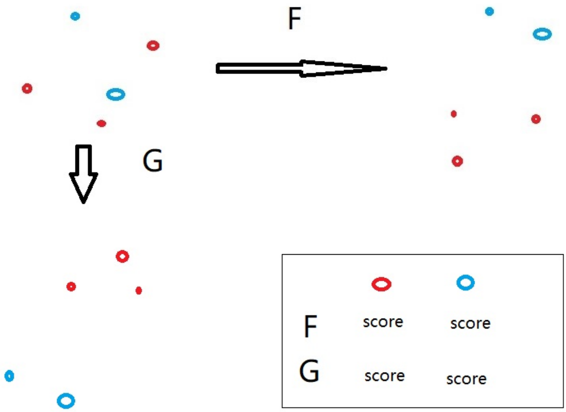
Machine Learning



Selecting Operators



Operators Sampling



Metric Learning

1. Select a dataset $D \in$ {MNIST, fashion-MNIST, CIFAR-10}.
2. Select two random classes l_1, l_2 among the classes labelling D and consider the sets Φ_1, Φ_2 of samples associated to the selected classes.
3. Preprocess the images according to the procedure described as above.
4. Choose randomly $n << |\Phi_i|$ samples per class ($|\Phi_i| = 6000$ for every i and $n \in \{20, 40\}$ in our experiments).
5. Initialize, select and sample IENEOs by evaluating them on the random samples extracted at the previous step.
6. Use the selected and sampled IENEOs to compute distances between validation samples (10 per class).

Metric Learning

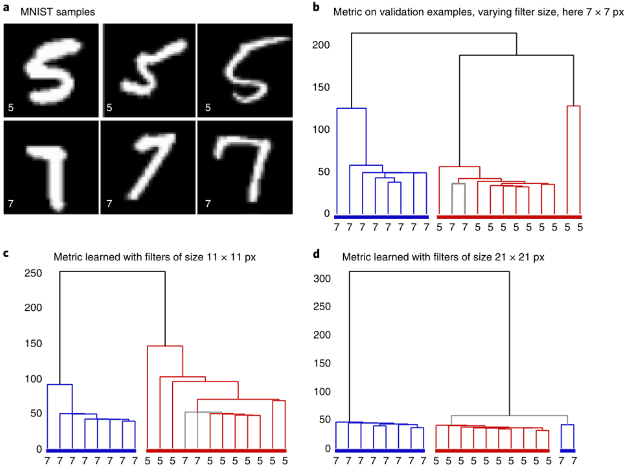


Figure 1: Metric learning on MNIST

Metric Learning

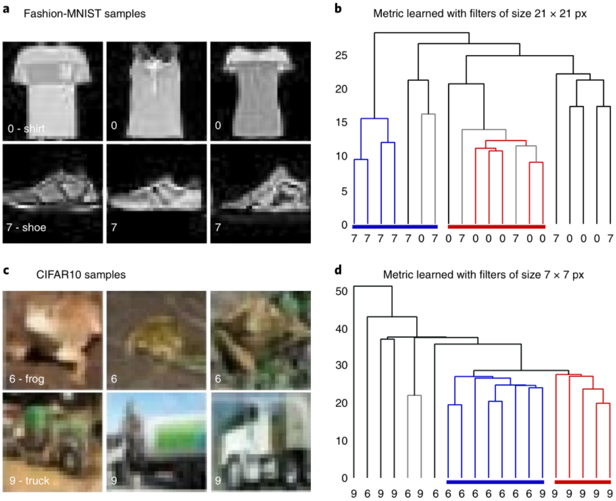


Figure 2: Metric learning on fashion-MNIST and CIFaAR-10

Advantages of GENOEs

- The space of GENOEs can be controlled by group actions.

$$\{F : (\Phi, G) \to (\Psi, G) | F(\phi \circ g) = F(\phi) \circ g, \forall g \in G\}.$$

- The learning method is not the usual weight adjustment, but choosing a minimal and effective space of operators by simpling.
- The size (other than the weight parameter) of convolution kernels can be learned.

Injecting Prior Knowledge

- First, we provide additional evidence that the filters obtained through the selection and sampling procedures are informative with respect to a chosen dataset, even when evaluated on few samples per class.
- Second, we aim to show that this information can be used for classification and hence injected as prior knowledge in an artificial neural network architecture under the form of fixed convolutional kernels.

Injecting Prior Knowledge

Dataset	500 operators	500 operators	750 operators	Random filters
	20 examples	40 examples	20 examples	
MNIST	0.988591	0.963282	**0.991158**	0.506815
MNIST	**0.989732**	0.949096	0.987450	0.960501
MNIST	**0.990588**	0.960501	0.989447	0.506815
f-MNIST	0.984598	0.960223	**0.986309**	0.506815
f-MNIST	0.990017	0.958275	**0.990588**	0.958554
f-MNIST	**0.990302**	0.963561	0.988591	0.506815
CIFAR-10	0.982886	0.960223	**0.987735**	0.506815
CIFAR-10	**0.992299**	0.962170	0.988876	0.506815
CIFAR-10	0.982316	0.960501	**0.989732**	0.971627
MNIST $\mu(\sigma)$	0.989(0.001)	0.958(0.007)	0.989(0.001)	0.658(0.262)
f-MNIST $\mu(\sigma)$	0.988(0.003)	0.961(0.002)	0.988(0.002)	0.657(0.261)
CIFAR-10 $\mu(\sigma)$	0.985(0.005)	0.961(0.001)	0.988(0.001)	0.661(0.268)
All $\mu(\sigma)$	0.988(0.003)	0.959(0.004)	0.988(0.001)	0.659(0.228)

Comparison between the performance obtained by classifying two classes of MNIST, fashion-MNIST (f-MNIST) and CIFAR-10 with a dense classifier fed with features extracted by convolutional filters obtained by selecting and sampling IENEOs. Column headers specify the dataset and the initialization parameters used to select and sample the IENEOs. The last column is the control for our experiments, where the classification is based on features extracted with random filters. Experiments have been repeated three times for each dataset and initialization parameters. In the columns, the accuracy obtained by the classifier on the validation set (3,000 images per class) is reported, along with the dataset-wise and overall mean accuracy and standard deviation. Per row best results are reported in bold numbers.

Figure 3: Evaluating the knowledge injection capability of GENEOs

IENEOs

IENEOs: a parametric family of non-expansive operators that are equivariant with respect to Euclidean plane isometries.

Given $\sigma > 0$ and $\tau \in \mathbb{R}$, we consider the Gaussian function with width σ and centre τ

$$g_\tau(t) := \exp\left(-\frac{(t-\tau)^2}{2\sigma^2}\right)$$

where $g_\tau : \mathbb{R} \to \mathbb{R}$. For a positive integer k, we take the set S of the $2k$-tuples $(a_1, \tau_1, \ldots, a_k, \tau_k) \in \mathbb{R}^{2k}$ for which $\sum a_i^2 = \sum \tau_i^2 = 1$. Then S is a submanifold of $\mathbb{R}^{2k}$.

IENEOs

For each $p = (a_1, \tau_1, \ldots, a_k, \tau_k) \in S$, we have a function

$$G_p(x, y) := \sum_{i=1}^{k} a_i g_{\tau_i}(\sqrt{x^2 + y^2}).$$

We denote by F_p the convolutional operator mapping each continuous function with compact support $\varphi : \mathbb{R}^2 \to \mathbb{R}$ to the continuous and compactly supported function $\psi : \mathbb{R}^2 \to \mathbb{R}$ defined as

$$\psi(x, y) := \int_{\mathbb{R}^2} \varphi(\alpha, \beta) \frac{G_p(x - \alpha, y - \beta)}{\|G_p\|_{L^1}} d\alpha d\beta.$$

IENEOs

- The operator F_p is a GENEO with respect to the group I of Euclidean plane isometries.
- F_p is an IENEO, and we denote $\mathcal{F} = \{F_p\}_{p \in S}$.
- It is similar to CNN by learning a family suitable operators (convolution kernels).

Image pre-processing

- Every image ι is first reshaped to size (128,128).
- Then it is blurred with a 3×3 Gaussian kernel.
- Finally, it is standardized as $\iota = \frac{\iota - \bar{\iota}}{\sigma(\iota)}$.

Operator selection and sampling on labelled datasets

1. Let Φ be a dataset labelled by $l : \Phi \to J \subset \mathbb{N}$. We assume that the dataset can be written as the disjoint union $\Phi = \bigsqcup_{i \in J} \Phi_i$.
2. Let $\mathcal{F}$ be the space of IENEOs, and choose randomly a set $\mathcal{C} = \{F_k\}_{k \in \{1,\dots,N\}}$.
3. Select those operators in $\mathcal{C}$ that consider as similar the objects belonging to the same class by the rule

$$s_l(F) < \varepsilon, \quad \forall l \in J,$$

for a fixed threshold ε, where

$$s_l(F) = \max_{\varphi_i^l, \varphi_j^l} d_{\mathrm{match}}(\beta_1(F(\varphi_i^l)), \beta_1(F(\varphi_j^l))).$$

Operator selection and sampling on labelled datasets

The selection criterion does not guarantee that the operators are maximally diverse.

1. Given a class l, we define the distance between two operators F_p and F_q as

$$\Delta^l_{GENEO}(F_p, F_q) := \sup_{\varphi^l_i \in \Phi_l} d_{\mathrm{match}}(\beta_1(F_p(\varphi^l_i)), \beta_1(F_q(\varphi^l_i))).$$

   We select only one operator for pairs whose score is below a fixed threshold t.

2. Two objects φ_1 and φ_2 can be compared by computing the strongly G-invariant pseudo-metric

$$\mathcal{D}^{\mathcal{F}}_{\mathrm{match}}(\varphi_1, \varphi_2) := \sup_{F \in \mathcal{F}} d_{\mathrm{match}}(\beta_1(F(\varphi_1)), \beta_1(F(\varphi_2))).$$

Metric learning through selection and sampling

Selected and sampled operators $F_i \in S$ can be used to measure distances between pairs of validation samples as

$$d_S(\varphi, \varphi') = \max_{F \in S} d_{\mathrm{match}}(\beta_1(F(\varphi)), \beta_1(F(\varphi'))).$$

Note that d_S is invariant with respect to the action of the group of planar isometries.

Thank You!