Group Equivariant CNN's

April 10, 2021

Classification of data

Definition

Let X be a compact topological space, let $\Phi = \{f : X \to \mathbb{R}^d \quad \textit{continuous}\}$ and $\mathcal{L} = \{1, ..., L\}$ a set of labels, we call *classification* a function

$$\mathcal{C} : \Phi \to \mathcal{L}.$$

This classification fuction defines an equivalence relation in Φ.

Definition

Let $f, \varphi \in \Phi$ we say that $f \stackrel{\mathcal{C}}{=} \varphi \stackrel{def}{\iff} \mathcal{C}(f) = \mathcal{C}(\varphi)$.

Classification of data

Every equivalence relation on Φ fully determines a symmetry group.

Definition

Let $\simeq$ be an equivalence relation on Φ, then the *symmetry group* defined by $\simeq$ is

$$G := \{g \in Homeo(X) : \forall f \in \Phi \quad f \circ g \simeq f\}.$$

Conversely, any group also defines an equivalence relation on Φ.

Definition

Let $f, \varphi \in \Phi$, let G be a subgroup of $Homeo(X)$,
we say that $f \overset{G}{=} \varphi \overset{def}{\iff} \exists g \in G : f \circ g = \varphi$.

Natural pseudodistance

Our understanding of data is based on the distinction of such data with respect to a symmetry group G, i.e. a distance on Φ.
Every data recognition task is therefore based on the pair (Φ, G) which we call *perception pair*.

Definition

Let (Φ, G) be a perception pair, we define the *natural pseudodistance* $d_G : \Phi \times \Phi \to \mathbb{R}$ to be

$$d_G(f, \varphi) := inf_{g \in G} |f - \varphi \circ g|_{\infty}.$$

While Φ is not Hausdorff with respect to d_G, we have that $[\Phi]_{\underline{\underline{G}}}$ is.

G-equivalence and $\mathcal{C}$-equivalence

The two equivalences are related if G is the symmetry group of Φ with respect to a classification $\mathcal{C}$

Proposition

Let $f, \varphi \in \Phi$ and let $G := \{g \in Homeo(X) : \forall f \in \Phi \quad f \circ g \overset{\mathcal{C}}{=} f\}$, then

$$f \overset{G}{=} \varphi \implies f \overset{\mathcal{C}}{=} \varphi.$$

The problem with this approach is that we need to know the symmetry group of our recognition task a priori. And that if two pieces of data are equally labeled by $\mathcal{C}$ we are not sure that there exists a transformation in $Homeo(X)$ that maps one into the other.

G-equivariant network

Let's start from a known symmetry of the data. In order to have a neural network able to classify objects, we need it to learn G-invariance and G-equivariance.

Definition(G-equivariant neural network)

A G-equivariant neural network is a sequence of operators $\{C_j : \Phi \to \Phi\}$ such that

$$\forall f \in \Phi, g \in G \quad \exists g' \in G : C_j R_g f = R_{g'} C_j f.$$

The G-equivariant operators preserve G-equivalences.

Proposition

Let $f, \varphi \in \Phi$, $\{C_j : \Phi \to \Phi\}$ a G-CNN, then $f \overset{G}{=} \varphi \implies C_j f \overset{G}{=} C_j \varphi$.

We want data equal in the first layers to be equal also in the following layers.

G-equivariant network

Nevertheless some distinguishable data will become indistinguishable in the following layer, this non-expansiveness is the key for abstraction. The last layer must obviously be G-invariant.

Let's see how to construct G-equivariant layers.

G-equivariant convolution

Definition(G-convolution first layer)

Let $f \in \Phi$, i.e. $f_i : X \to \mathbb{R} \quad i \in I = \{1, ..., d\}$, let $\psi^i_j : X \to \mathbb{R} \quad j \in I, i = 1, ..., n$ be the filters, then the first convolutional layer is

$$[f * \psi^i](g) = \sum_{x \in X} \sum_{j \in I} f_j(x) \psi^i_j(g^{-1}x).$$

Definition(G-convolution hiddel layers)

Let $f_i : G \to \mathbb{R} \quad i \in I = \{1, ..., n\}$, let $\psi^i_j : G \to \mathbb{R} \quad j \in I, i = 1, ..., m$ be the filters, then the first convolutional layer is

$$[f * \psi^i](g) = \sum_{h \in G} \sum_{j \in I} f_j(h) \psi^i_j(g^{-1}h).$$

G-equivariant convolution

Proposition

G-convolution is G-equivariant.

Similarly also pooling and non-linearities are.

Graph Neural Networks

Let C_0 be the space of 0-chains, and C_1 the space of 1-chains of a graph. The graph layers are defined as low degree polynomials of local aggregation functions.

Definition

We call a local aggregation function a linear function $L : C_0 \to C_0$ such that

$$L_{ij} = M_{ij} \odot A_{ij},$$

where A_{ij} is the adjacency matrix and M_{ij} is any matrix.

Definition

We call *diffusion group* $\mathcal{D}_{\mathcal{A}}$ w.r.t. $A : C_0 \to C_0$, the image of the $(\mathbb{R}, +)$ homomorphism

$$t \mapsto e^{At} : C_0 \to C_0.$$

Diffusion Law

Definition

We call *diffusion law* w.r.t. $A : C_0 \to C_0$ the equation

$$\frac{d}{dt}|f(t)\rangle = -A|f(t)\rangle.$$

Proposition

The solution of the diffusion equation is

$$|f(t)\rangle = e^{-At}|f(0)\rangle.$$

Proposition

Let $pol(A)$ be any polynomial of A, then

$$[pol(A), e^{-At}] = 0.$$

Conclusion

Since the diffusion law is learnable, we have that the GNN learns a diffusion law for any feature. And a diffusion equivariant layers do not distinguish between a chain and its diffusion at a time t.

Proposition

$|f\rangle \overset{\mathcal{D}_{\mathcal{A}}}{=} |g\rangle$ if g is a diffusion of f.