DomainBed

起因

当前的机器学习系统在面对新的示例分布时会反复无常地失败，阻碍了这项技术不能在关键领域的使用.

我们希望找到鲁棒的机器学习模型，能够规避假相关。让模型发现不因域内域外而改变的模式

结论

- ERM (Empirical risk minimization) Baseline优于其他结果，主要原因有：
    - 更大的网络架构（ResNet-50）
    - 强大的数据增强
    - 仔细的参数调优
- 强数据增强可以提高分布外泛化，同时不影响分布内泛化
- 消除虚假相关性

评价

各个数据集上的精度不加权重的取均值不太合理。导致所有模型的结果都在66%左右。

一个有趣的现象是，所有模型在CMNIST上的结果都不高，50%左右。据我了解这是份加了颜色的手写字数据集，不是很复杂，精度应该比较高才对。值得深究的一个点。

further

CMNIST值得看一下

此论文中实验部分有必要详细看一下，看下他怎么拿数据集进行训练的。以至于CMNIST上精度那么低。

Risk minimization

Paper title: Principles of Risk Minimization

for Learning Theory
Learning is posed as a problem of function estimation, for which two principles of solution are considered: **empirical risk minimization** and **structural risk minimization**. These two principles are applied to two different statements of the function estimation problem: **global and local**. Systematic improvements in prediction power are illustrated in application to zip-code recognition.

相机内参外参

内参

$$Z\begin{pmatrix}u\\v\\1\end{pmatrix}=\begin{pmatrix}f_x & 0 & c_x\\0 & f_y & c_y\\0 & 0 & 1\end{pmatrix}\begin{pmatrix}X\\Y\\Z\end{pmatrix}=KP$$

图像坐标系
K内参矩阵，出厂后一般就固定了.
P相机坐标系下的坐标，以相机为原点.

外参

世界坐标系：在环境中选择一参考坐标系来描述相机的位置，该坐标系为世界坐标系。又叫系统的绝对坐标系。

相机位姿由旋转矩阵R，平移向量t描述，为相机的**外参**。

$$P=RP_W+t$$

相机的外参决定了相机的位姿.

似乎，旋转矩阵R是正交矩阵,$R^T = R^{-1}$

Quaternions 四元数

四元数的优势在于表示旋转.
Quaternions are cool. Even if you don't want to use them, you might need to defend yourself from quaternion fanatics.

Why can't we invert vectors in $\mathbb{R}^3$?

$\mathbb{R}^1, \mathbb{R}^2$中元素可以求逆（复数），但是$\mathbb{R}^3$不能.

$$q_0 + q_1 i + q_2 j + q_3 k$$

$$i^2 = j^2 = k^2 = ijk = -1$$

共轭： q^*

$qq^* = |q|^2$

$$q^{-1} = \frac{q^*}{|q|^2}$$

单位四元数$|q| = 1$，可以表示为

$$q = \cos\frac{\theta}{2} + \sin\frac{\theta}{2}\hat{\mathbf{n}}$$

$(q^{-1} = q^*)$

令$x = 0 + \mathbf{x}$为一纯四元数，
令$x' = qxq^*$
则x'是x绕$\hat{\mathbf{n}}$旋转θ角后的一个纯四元数.

这是单位四元数的几何意义

Remark

给一四元数，可以得到旋转轴和旋转角，以及旋转矩阵.

Pitch, yaw, roll

There are in fact six degrees of freedom of a rigid body moving in three-dimensional space.

forward/back, up/down, left/right, pitch, yaw, roll.

- pitch: nose up or tail up.
- yaw: nose moves from side to side.
- roll: a circular (clockwise or anticlockwise) movement of the body as it moves forward.

nuscenes

Sensor synchronization

摄像机一直开着，这很费电的啊。

为了在激光雷达和相机之间实现良好的跨模态数据对齐，当顶部激光雷达扫过相机视场中心时，会触发相机的曝光。图像的时间戳为曝光触发时间;激光雷达扫描的时间戳为当前激光雷达帧实现全旋转的时间。鉴于相机的曝光时间几乎是瞬时的，这种方法通常会产生良好的数据对齐。**请注意，摄像机以12Hz运行，而激光雷达以20Hz运行。12个相机曝光尽可能均匀地分布在20个激光雷达扫描中，因此并非所有激光雷达扫描都有相应的相机帧**。将摄像机的帧率降低到12Hz有助于降低感知系统的计算、带宽和存储需求。

AE-OT generator

值得注意的点

- DNNs 只能表示连续映射
- 生成器是传输映射，传输白噪音因分布到数据集分布.通常这种映射是不连续的.
- We propose that these phenomena relates deeply with the singularities of distribution transport maps.（我们认为这些现象与分布输运图映射的奇异性密切相关。）
- 流形分布假设在深度学习中被广泛接受，它假设特定类别的自然数据的分布集中在嵌入在高维数据空间中的低维流形上。
- 因此gan和vae隐含的目标是完成两个主要任务:
    - 流形嵌入:寻找嵌入到图像空间的数据流形与潜在空间之间的编码/解码映射;（to find the encoding/decoding maps between the data manifold embedded in the image space and the latent space;）
    - 概率分布传输:将给定的白噪声分布传输到数据分布中，可以传输到潜在空间中，也可以传输到图像空间中。probability distribution transport: to transport a given white noise distribution to the data distribution, either in the latent or in the image space.

做法

将映射拆成两部分：流形嵌入+最优传输.

理论

- 模式坍塌原因，概括一下，目标集**非凸**.
    - 问题：非凸会带来什么影响？凸集又有哪些优势？
- (single mode)
    - Brenier’s polar factorization theorem (1991)
    - Figalli’s regularity theorem (2010)

语句

在贡献结尾写这一条很不错，(iv) Our experiment results demonstrate the efficiency and efficacy of the proposed method.

最优传输

Brenier interview

- How do you discover the link between optimal transport and hydrodynamics?
- 你对年轻人的建议？
    - To keep some strong personality, avoid to be much follower. Okay, it probably may be a wrong advice. Because it is the most profitable thing to do, to be in a good field, with very good leaders. But i would ... to be free as much as possible. To try to open you our track research.

Brenier lecture: The melting rubik cube: From Fluids to Combinatorics and vice versa.

- 在流体力学中，欧拉是许多著名科学家的跟随者。但是他是第一个，明确地描述流体的人（1755，用偏微分方程）。
- 不可压缩的流体，被限制在区域D中，并且按照欧拉方程流动。just follows a (constant speed) geodesic curve along the manifold of all possible incompressible maps of D.

一个打乱的魔方，变成液体后流动之后，再还原，变成一个未打乱的模仿。似乎在讲这个意思。

顾险峰老师课件

- Monge问题求最优映射，kantorovich问题是弱化为最优方案.
- 如果最优传输映射存在的话，不要用线性规划去求解。很多0，造成浪费。

Monge问题与Kantorovich问题

Monge问题

$$(MP) \quad \inf\left\{M(T) := \int_X c(x, T(x)) d\mu(x) :\ T_{\#}\mu = \nu\right\}$$

note:

$$\nu(A) = \mu(T^{-1}(A))$$

Kantorovich问题连续形式

$$\min_{\gamma} \int_{X\times Y} c(x, y) d\gamma(x, y)$$

相对于γ是一个线性泛函

约束为无限维的凸约束

广义Lagrange乘子法转为对偶问题

广义：约束无限个

ϕ, ψ可以看作Lagrange乘子或者影子价格.

公式演义

惩罚项：

$$\begin{aligned}
&\forall \gamma \in \mathcal{M}_+(X\times Y) \\
&\qquad\qquad\qquad\qquad \sup_{\phi,\psi} \int_X \phi d\mu + \int_Y \psi d\nu - \int_{X\times Y} (\phi(x) + \psi(y)) d\gamma \\
&\qquad\qquad\qquad\qquad = \begin{cases} 0, \gamma \in \Pi(X, Y) \\ +\infty, \gamma \notin \Pi(X, Y) \end{cases}
\end{aligned}$$

代入Kantorovich问题

Kantorovich原问题（离散形式）

$$\min_{\gamma} \sum_{i,j}^{m,n} c_{ij}\gamma_{ij}$$

$$s.t. \begin{cases} \sum_j \gamma_{ij} \geq \mu_i \\ \sum_i \gamma_{ij} \geq \nu_j \end{cases}$$

Kantorovich对偶问题（离散形式）

$$\max_{\phi,\psi} \sum_i^m \phi_i\mu_i + \sum_j^n \psi_j\nu_j$$

$$s.t. \begin{cases} \phi_i + \psi_j \leq c_{ij} \\ \phi_i, \psi_j \geq 0 \end{cases}$$

Weierstrass 定理

f下半连续，X紧，则存在$\bar{x} \in X$,满足

$$f(\bar{x}) = \min\{f(x) : x \in X\}$$

Kantorovich问题解的存在性

- 紧空间连续代价函数
- 紧空间下半连续代价函数
- Polish空间（完备的可分度量空间）下半连续代价函数

逐步推广，工程上只用第一种.

Brenier理论 对计算很有用，和微分几何也有联系

什么时候最优传输方案一定是最优传输映射

$X, Y \subset \mathbb{R}^d$是欧式空间子集，$c(x, y) = h(x - y), h : \Omega \to \mathbb{R}$是一个严格凸函数. 这时最优传输方案一定是最优传输映射，并可以用公式直接表达.

引理

γ是最优传输方案，(x_0, y_0)属于γ的支撑集，有

$$\bigtriangledown_x c(x_0, y_0) = \bigtriangledown\varphi(x_0)$$

支撑和包络在这一点相切的解析表达.

扭曲条件（充分条件）

- c关于x处处可微
- $\forall x_0, y \mapsto \bigtriangledown_x c(x_0, y)$是单射

方案是映射的充分条件

凸函数的梯度映射

h是一个C^2严格凸函数，定义域是凸集，则梯度映射$x \mapsto \bigtriangledown h(x)$全局可逆，并且是微分同胚，记为$(\bigtriangledown h)^{-1}$

定理

$$T(x) = x - (\bigtriangledown h)^{-1}(\bigtriangledown\varphi(x))$$

势能的微分给出了映射

Brenier定理

当$c(x, y) = \frac{1}{2}|x - y|^2$时，

$$T(x) = x - \bigtriangledown\varphi(x)$$

$$T(x) = \bigtriangledown u(x)$$

$$u(x) = \frac{1}{2}|x|^2 - \varphi(x)$$

是凸的

Monge-Amp$\grave{e}$re方程

$(\Omega, \mu), (\Omega^*, \nu), \Omega, \Omega^* \subset \mathbb{R}^d$是凸紧集，
$d\mu(x) = f(x)dx, d\nu(y) = g(y)dy$连续，$u$是$C^2$光滑，$T$保测度，有

$$\det D^2 u(x) = \frac{f(x)}{g \circ \bigtriangledown u(x)}$$

具有第二边界条件

$$\bigtriangledown u(\Omega) = \Omega^*$$

计算方法

FFT-OT算法

最近才发表，内容比较新。
计算最优传输方案主要是用Kantorovich发明的线性规划，和它的各种各样的变形，最主要是sinkhorn方法

俄罗斯学派用光滑函数逼近分段线性函数是目前比较热的一种方法

本质上都是在解Monge-Amp$\grave{e}$re方程

这个方程是强烈非线性的，传统方法很难解

其实就是对PDE中的一个非线性算子，我们局部线性化.

步长不能太大 所以要在可容许的空间中进行搜索，否则整个方法就崩溃.

方程的正则性理论 很重要.

这些解的收敛速度，收敛状态是开放问题，可以做计算数学方向的博士论文。

二维情形——不动点方法

$$u_{xx}u_{yy} - u_{xy}^2 = f/g \circ Du$$

转换为Poisson方程

$$\triangle u = \sqrt{u_{xx}^2 + u_{yy}^2 + 2u_{xy}^2 + 2f/g \circ Du}$$

$\triangle = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$

定义算子

$$\mathcal{T} : H^2(\Omega) \to H^2(\Omega)$$

$$\mathcal{T}[u] = \triangle^{-1}\left\{\sqrt{u_{xx}^2 + u_{yy}^2 + 2u_{xy}^2 + 2f/g \circ Du}\right\}$$

则
方程的解是算子$\mathcal{T}$的不动点.
可以用迭代法得到不动点

$$u^{(n+1)} = \mathcal{T}[u^{(n)}]$$

.

Neumann边界条件

换元
令$u(x,y) = \varphi(x,y) + (x^2+y^2)/2$
得

$$Du = D\varphi + Id$$

$$\triangle u = \triangle\varphi + 2$$

算子$\mathcal{T}[u]$被变换成了$\mathcal{P}[\varphi]$

$$\mathcal{P}[\varphi^{(n+1)}] := \triangle^{-1}\mathcal{F}[\varphi^{(n)}]$$

这里

$$\mathcal{F}(\varphi) := \left\{\sqrt{(\varphi_{xx}+1)^2 + (\varphi_{yy}+1)^2 + 2\varphi_{xy}^2 + 2f/g \circ (Id + D\varphi)} - 2\right\}$$

边界条件

$$\partial\varphi/\partial\mathbf{n} = 0.$$

每个点和像点之差与边界垂直

疑问：D和$\bigtriangledown$是不是一个意思？

有限差分法

差分法的好处是可以用硬件加速.

有限差分算子

$$\mathcal{D}_{xx}^2 u_{ij} = \frac{1}{h_x^2}(u_{i+1,j} + u_{i-1,j} - 2u_{ij})$$

$$\mathcal{D}_{yy}^2 u_{ij} = \frac{1}{h_y^2}(u_{i,j+1} + u_{i,j-1} - 2u_{ij})$$

$$\mathcal{D}_{xy}^2 u_{ij} = \frac{1}{4h_x h_y}(u_{i+1,j+1} + u_{i-1,j-1} - u_{i-1,j+1} - u_{i+1,j-1})$$

离散Poisson方程

$$u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{ij} = \rho_{ij}$$

离散余弦变换DCT

DCT DST（狄利克雷边界条件）
给定二维数组$u(i,j)$，二维DCT定义为

$$\tilde{u}(m,n) = c(m,n)\sum_{i,j} u(i,j)\cos\frac{(2i+1)m\pi}{2M}\cos\frac{(2j+1)n\pi}{2N}$$

$$m,i = 0,1,\cdots,M-1; \quad n,j = 0,1,\cdots,N-1$$

$$c(m,n) = \begin{cases} \frac{\sqrt{2}}{\sqrt{MN}}, m=0, n=0 \\ \frac{2}{\sqrt{MN}}, else \end{cases}$$

在频域，Poisson方程的解可以直接写出来
引理

给定离散Poisson方程，具有Neumann边界条件：

$$\triangle u=\rho, \frac{\partial u}{\partial \mathbf{n}}=0$$

令

$$\tilde{\rho}=DCT(\rho), \quad \tilde{u}=DCT(u)$$

有

$$\tilde{u}(m,n)=\frac{\rho(\tilde{m},n)}{2[\cos\frac{m\pi}{M}+\cos\frac{n\pi}{N}-2]}$$

不同相差一个常数，令$\tilde{u}(0,0)=0$得到唯一解.
Opencv 做图像压缩，第一步通常是fft,ifft。
把三维曲面映射到二维平面，理论上最严密的是共形映射
局部特征不变

这种方法在高维的应用不是很多

矩阵行列式的线性化（线性化Monge-Ampère方程）

单位矩阵邻域内行列式的线性化公式

$$\det(I+\varepsilon N)=1+\varepsilon\cdot tr[N]+O(\varepsilon^2)$$

我们寻找矩阵M附近行列式的线性化

$$\begin{aligned}\det(M+\varepsilon N)&=\det(M)\det(I+\varepsilon M^{-1}N)\\&=\det(M)\cdot(1+\varepsilon\cdot tr[M^{-1}N]+O(\varepsilon^2))\\&=\det(M)+\varepsilon\cdot tr[\det(M)M^{-1}N]+O(\varepsilon^2)\\&=\det(M)+\varepsilon\cdot tr[M_{adj}N]+O(\varepsilon^2)\end{aligned}$$

这里$M_{adj}:=\det(M)M^{-1}$（伴随矩阵）

行列式的线性化算子表示为：

$$\bigtriangledown_M \det(M)[N] := tr(M_{adj}N)$$

线性化Monge-Ampère方程算子

$$\det D^2 u(x) = \frac{f(x)}{g \circ \bigtriangledown u(x)}$$

年轻的时候拓宽视野.

跨领域成为专家还是很难的.

所以多学点基础数学，越靠近基本的话，以后的天地越大.

所以机器学习兴起之后，年轻人的知识结构产生了巨大的断层.

当深度学习热潮褪去之后，这批学者在学术界做经典的研究就会比较吃力.