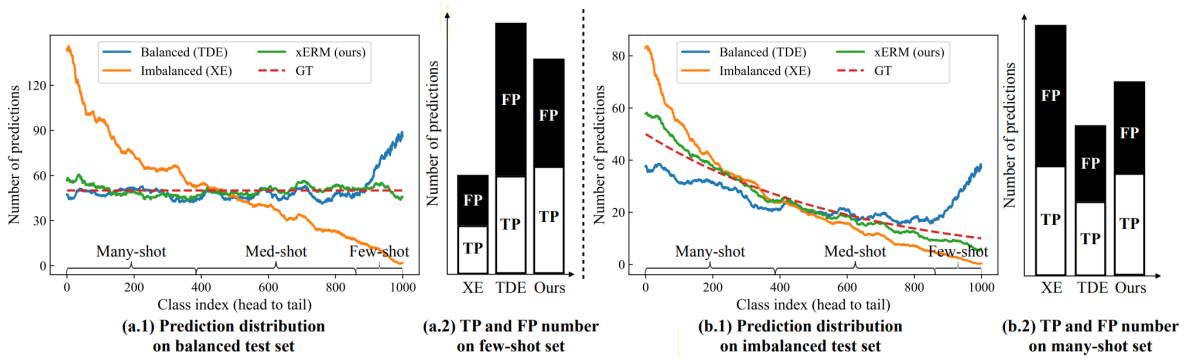
Cross-Domain Empirical Risk Minimization for Unbiased Long-tailed Classification

这篇paper还是很值的一读的，里面思路和我自己最近在做得十分像，里面从分析到最后的模型都和我想得很像，不过论文分析的很到位，我觉得还是很有启发的（之前读了作者他们TDE那篇paper，后来觉得TDE那篇paper最后model不是特别合理，这次这篇文章也指出来了这些问题）。

Introduction

在分析部分，总得来说论文觉得过去的那些方法事实上都没有真正解决长尾问题中的类别偏见，最开始是偏向head类，而用了长尾问题的一些trick后变成偏向tail类别，论文给出了证据（其中这种bias其实意味着对应的FP比较多）：



(a.1) Prediction distribution on balanced test set　(a.2) TP and FP number on few-shot set　(b.1) Prediction distribution on imbalanced test set　(b.2) TP and FP number on many-shot set

至于为什么尽管用了长尾trick后头部类性能还可以，论文猜想是因为一方面头部类样本足够多，另一方面测试集是均衡的。

所以说目前长尾问题还没有从本质上得到解决，要去很好的在两种数据分布之间找到一个平衡是十分困难的。

Related Work

- 第一块长尾分布的在其它笔记里有，都差不多。
- 第二块因果分析的，论文提到文中方法是另外一篇因果分析论文方法的一种应用，没去看过这篇引用paper。
- 第三块是关于bias migration的，论文提到一篇和本文有关系得模型LfF。

Method

从方法上看论文提出的方法不复杂挺直接的。

- 第一步，确定model。

就是找两个model（一个也行分成两部分），一个代表imbalance（会偏向头部类），另一个代表balance（实际这种平衡指的是不偏向于头部类，一般之前的方法都是偏向尾部类，上面introduction里有说）。

这两个model可以替换，后面论文是用了之前他们组提出的的TDE还有一个PC，两个都是后验调整logit的，这两个方法本身就属于balance model，把这两个model的后验调整那块去掉就是imbalance model了，这样用一个model分为两个部分就行，计算开销比较小。

- 第二步，合并两个model。

两个model合并起来的方式就是把loss加权起来，那只要知道权重就好了，权重的话就是根据CE loss的大小，imalance和balance两部分，谁的CE loss大一点谁权重就大一点。有一点要注意一下，这里的合并并不是说两个model的结果都作为参考依据，实际上是模拟了一个采样的过程，这里的CE loss并不是后面用的loss。

下面这个loss第一部分就是原来的CE，而第二部分的R^{ba}在下一张图片里有。

$$\mathcal{R}(f) = \mathcal{R}^{\text{imba}}(f) + \mathcal{R}^{\text{ba}}(f).$$

$$\text{Balanced Domain ER:} \qquad \mathcal{R}^{\text{ba}}(f) = -w^{\text{ba}} \sum_i \hat{y}_i \log f_i(x), \tag{4}$$

where $\hat{y}_i = p^{\text{ba}}(y_i|x)$ denotes the balanced prediction for i-th class. The overall empirical risk minimization:

要注意一下的就是图片里这个balance的loss计算方法，这个后面会拿因果理论证明，有点伪标签的意思。

- 具体从motivation来看下前面两步。

前面method有三个问题，第一，为什么要用这样一个model；第二，那个balance loss为什么这样计算；第三，最后那个combine的权重为什么那样定。这些问题paper用因果分析来解释。

事实上，第一步为什么要引入balance和不balance，这是因为作者认为balance和不balance这个因素暗中影响了由x预测到y这一步，这里其实看一下我之前TDE那个ppt就可以了，原理一样，同时可以得到下面这个loss公式，do(x)就表示了是属于balance方法的还是imbalance方法的。

$$\begin{aligned}\mathcal{R}(f) &= \mathbb{E}_{x\sim P(X), y\sim P(Y|do(X=x))}\mathcal{L}(y, f(x)) \\ &= \sum_x \sum_y \mathcal{L}(y, f(x)) P(y|do(x)) P(x),\end{aligned}$$

里面就中间这个P(y|do(x))需要求一下，也就是按照是否属于balance分布的：

$$\begin{aligned}P(y|do(x)) &= \sum_{S=s\in\{0,1\}} P(y|x, S=s)P(S=s) \\ &= \frac{P(x, y, S=1)}{P(x|S=1)} + \frac{P(x, y, S=0)}{P(x|S=0)}.\end{aligned} \tag{7}$$

上面两个公式合在一起就是，最后那个P(x,y,s)变成了$1/N$：

$$\begin{aligned}\mathcal{R}(f) &= \sum_{(x,y)} \sum_{s\in\{0,1\}} \mathcal{L}(y_s, f(x)) \frac{P(x)}{P(x|S=s)} P(x,y,s) \\ &= \frac{1}{N} \sum_{(x,y)} [\underbrace{\mathcal{L}(y_{s=1}, f(x)) \frac{P(x)}{P(x|S=1)}}_{\text{Imbalanced Domain ER}} \\ &\quad + \underbrace{\mathcal{L}(y_{s=0}, f(x)) \frac{P(x)}{P(x|S=0)}}_{\text{Balanced Domain ER}}],\end{aligned} \tag{8}$$

现在就是loss之外的两个权重问题了，因为P(x)/P(x|s)=P(s)/P(s/x)，如果假设P(s=0)=P(s=1)，那么P(s)就可以不考虑了，然后就是：

$$\frac{P(x)}{P(x|s)} \propto \frac{1}{P(s|x)}$$

这里就是说，给定样本，权重是和它属于balance或者imbalance的可能性的倒数成正比的。拿imbalance为例，也就是说这个样本如果被判断越属于imbalance，那么它的权重就越低，那想一想CE loss，如果用一个imbalance模型预测时候它的CE loss比较高，是不是就说明这个样本不属于imbalance分布的呢，那么他的w^{imba}是不是就高了呢。总之，论文就是用CE loss辅助确认了一下权重。

$$\begin{aligned} w^{\text{imba}} &\propto \frac{1}{P(S=1|x)} \propto (XE^{\text{imba}})^{\gamma} \\ w^{\text{ba}} &\propto \frac{1}{P(S=0|x)} \propto (XE^{\text{ba}})^{\gamma}. \end{aligned}$$