

PRML 第三次作业

GRF 及 LLGC 图分类

学生： 槐 泽 鹏

导师： X X X

学号:xxxxxxxxxxxxxxxxxxxx

联系方式: xxxxxxxxxxxxxxx

2020 年 12 月 11 日

目 录

第 1 节 GRF	1
第 2 节 LLGC	2
第 3 节 实验	3
3.1 数据集	3
3.2 结果	4
3.3 分析	6
第 4 节 小结	8
第 5 节 参考文献	9
第 6 节 代码	9

第 1 节 GRF

GRF 全称为基于高斯随机场 (Gaussian Random Field) 的半监督分类^[2], 是 Zhu 等人于 2003 年发表在 ICML 上的一种基于图上构建高斯随机场进行图像分类的方法。下面简要介绍 GRF 的原理和过程。

假设现在有 l 个带标签样本, 记为 $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)$, 和 u 个不带标签样本, 记为 $(x_{l+1}, y_{l+1}), \dots, (x_{l+u}, y_{l+u})$ 。通常情况下, $l \ll u$ 。记 $n = l + u$ 为所有样本个数。下面将这 n 个样本统一构建为无向图的形式。

图数据, 最基本的一些定义有: 连接矩阵 W 、度矩阵 D 、拉普拉斯矩阵 L 、归一化的拉普拉斯矩阵 \hat{L} 。GRF 只涉及前两个矩阵。

(1) 构建连接矩阵 W

连接矩阵的第 i 行第 j 列代表第 i 个样本和第 j 个样本的连接权重。连接权重有多种形式, 可以根据是否连接将其设置为 0 1; 也可以根据规则设置为连续值。这里 GRF 根据高斯函数设置权重, 如 1 所示, 并且由于 GRF 采用的是无向图, 因此连接矩阵是对称矩阵。

$$W_{ij} = \exp\left(-\sum_{d=1}^m \frac{(x_{id} - x_{jd})^2}{\sigma_d^2}\right) \quad (1)$$

式 1 中 m 是指每个样本有 m 维, 比如对于本次大作业要采用的 MNIST 数据集, 其图像为 $28 * 28$ 维, 则 $m = 28 * 28 = 784$ 。 σ_d 是代表第 d 维距离长度的超参数。值得一提的是, 这一超参数十分影响后续分类精度。

(2) 构建度矩阵 D

$$\begin{aligned} D &= \text{diag} D_{ii} \\ D_{ii} &= \sum_{j=1}^n W_{ij} \end{aligned} \quad (2)$$

GRF 旨在通过权重在图上传播标签, 传播准则是将最小化以下能量函数。

$$E(f) = \frac{1}{2} \sum_{i,j} w_{ij} (f(i) - f(j))^2 \quad (3)$$

式 3 中 $f(i)$ 代表第 i 个样本的标签。近似的, 可以认为该能量函数旨在使得两个样本的权重逼近他们的“欧氏距离”。

经过推导, 最后 u 个不带标签样本的预测标签为

$$f_u = (D_{uu} - W_{uu})^{-1} W_{ul} f_l \quad (4)$$

式 4 中, f_l 是 l 个有标签样本的标签, D_{uu} 、 W_{uu} 和 W_{ul} 来源于矩阵分块, 如 5 所示。

$$\begin{aligned} W &= \begin{bmatrix} W_{ll} & W_{lu} \\ W_{ul} & W_{uu} \end{bmatrix} \\ D &= \begin{bmatrix} D_{ll} & \mathbf{0} \\ \mathbf{0} & D_{uu} \end{bmatrix} \end{aligned} \quad (5)$$

第 2 节 LLGC

LLGC 全称为基于图局部和全局一致性 (Local and Global Consistency) 的高斯随机场的半监督分类^[1], 是 Zhou 等人于 2003 年发表在 Nips 上的一种基于图上临近节点构建高斯随机场进行图像分类的方法。LLGC 可以认为是基于 GRF 的改进方法, 主要在两点改进, 下面进行介绍。

(1) 能量函数中加入归一化处理

GGF 的能量函数 E 如式 3 所示, LLGC 在此基础上加入了归一化处理, 即使得不同样本之间更加注重角度距离而非绝对距离, 如 6 所示。

$$E_{LLGC}(f) = \frac{1}{2} \sum_{i,j} w_{ij} \left(\frac{f(i)}{\sqrt{d_i}} - \frac{f(j)}{\sqrt{d_j}} \right)^2 \quad (6)$$

(2) 能量函数中加入了标签正则项

具体地说, 就是标签在传播之后 (如 GRF 中的公式 4), 我们期望之前没有标签的样本获得新样本, 但是之前已有标签的样本尽可能少的改变其原本真实标签。因此加入标签正则项的新 LLGC 能量函数如 7 所示。

$$\begin{aligned} E_{LLGC}(f) &= \frac{1}{2} \sum_{i,j} w_{ij} \left(\frac{f(i)}{\sqrt{d_i}} - \frac{f(j)}{\sqrt{d_j}} \right)^2 \\ &\quad + \lambda \sum_{i=1}^n (f_i - y_i)^2 \end{aligned} \quad (7)$$

式 7 中, λ 为标签正则项系数超参数, f 为预测标签, y 为真实标签, 其中之前没有标签的样本这里全设置为 0。这里值得一提的是, 按照作者本来意图, 这一项不应该计算未标签样本, 即标签在图上传播之后, 只加入对已有标签样本的正则约束就可以, 但是这里对所有样本进行标签正则约束, 是因为如果不这么做后面就得不到解析解了。

以上两个部分即是 LLGC 相对于 GRF 改进的两个点。同样的, 在经过一番推导之后 (原文是一个迭代收敛的过程, 给出了无论初值是什么, 最终都能收

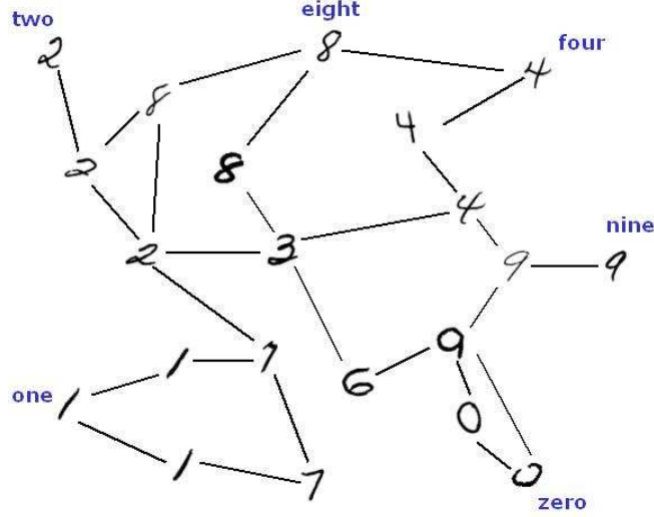


图 2.1 图数据建模示意图

敛至解析解的证明过程，我们这里直接使用其解析解即可)，标签更新公式为

$$F^* = (1 - \alpha)(I - \alpha S)^{-1}Y \quad (8)$$

式 8 中， α 为 $\frac{1}{1+\lambda}$ 是由严格推导得到的系数，与能量函数中标签正则项系数有关。 S 为用度矩阵归一化之后的连接矩阵，如 9 所示。 Y 为原始标签矩阵，值得注意的是，它的维度为 $n * c$ ， n 是总样本个数 (含未标签样本)， c 为标签类别个数， Y 矩阵的每一行类似于 one-hot 编码，属于哪一类则对应位置为 1，否则为 0；对于未标签样本，全置 0 (对应能量函数 7 中第二项求和时上角标为 n 而不为 l)。

$$S = D^{-1/2}WD^{1/2} \quad (9)$$

以上两节就是 GRF 和 LLGC 的全过程，最后附上原文中，将每个图象当作一个点，再构造为图数据的一个示意图，如图 2.1 所示。

第 3 节 实验

本节我们在 3 个数据集测试 GRF 和 LLGC 的效果。

3.1 数据集

一共采用 3 个数据集：MNIST、Vowel、Sonar，他们的静态参数如表 1 所示。

在超参数设置上，我们以 MNIST 数据集为例。我们将 σ 和 λ 分别设置为 (256, 32, 1024)、(0.1, 0.1, 0.9)，(256, 32, 1024) 是指从 256 开始，步长 32，到 1024 停止，共 24 个 (不含 1024)，(0.1, 0.1, 0.9) 同理，共 9 个 (含 0.9)。之后再行

循环跑参数测试，会得到共 $24 \times 9 = 216$ 对超参和对应的 GRF 和 LLGC 分类结果。这里将 σ 设置起步 256，这是因为 MNIST 数据集为图像，图像像素最大为 255，因此我们从 256 旨在使得连接矩阵中元素均小于 1。Vowel 和 Sonar 根据自身数据集特点有相似设置，详见表 2。

同时半监督中训练集比例设置为 0.2、0.3、0.4 三种。

以上两个超参数和训练集比例如表 2 所示。

表 1: 数据集参数

数据集名称	参数		数据本地位置及 网上下载链接
	标签类别数量	总样本数量 (训练集 + 测试集)	
MNIST	10	2000	train-images.idx3 train-labels.idx1 ¹
Vowel	11	990	vowel-contest.data ²
Sonar	2	208	sonar.all-data ³
备注： MNIST 的 2000 个数据集， 是从原始数据集的训练集 (6000 个) 中每类随机筛选 200 个， 最后总共 2000 个，再按照比例分割出训练集和测试集。			

表 2: 不同数据集的超参数

数据集名称	GRF	LLGC		训练集比例
	σ	σ	λ	
MNIST	(256, 32, 1024)	(256, 32, 1024)	(0.1, 0.1, 0.9)	0.2
Vowel	(0.02, 0.02, 10)	(0.02, 0.02, 10)	(0.1, 0.1, 0.9)	0.3
Sonar	(1/15, 1/15, 20/15)	(1/15, 1/15, 20/15)	(0.1, 0.1, 0.9)	0.4

3.2 结果

对 GRF 和 LLGC 的测试结果如 3 所示。

¹<http://yann.lecun.com/exdb/mnist/>

²<http://archive.ics.uci.edu/ml/machine-learning-databases/undocumented/connectionist-bench/vowel/>

³<http://archive.ics.uci.edu/ml/machine-learning-databases/undocumented/connectionist-bench/sonar/>

表 3: GRF 和 LLGC 分类结果

数据集	GRF		LLGC		训练集比例
	分类正确率	最佳超参 σ	分类正确率	最佳超参 $\sigma \lambda$	
MNIST	0.7769	384	0.8513	576, 0.1	0.2
	0.8057	384	0.8871	608, 0.1	0.3
	0.8125	352	0.9025	608, 0.1	0.4
Vowel	0.8283	0.2	0.8611	0.4, 0.1	0.2
	0.9033	0.2	0.9105	0.2, 0.1	0.3
	0.9276	0.2	0.9377	0.2, 0.1	0.4
Sonar	0.7470	0.2667	0.7349	0.6, 0.9	0.2
	0.8552	0.4667	0.8483	0.4667, 0.5	0.3
	0.728	0.2667	0.72	0.2667, 0.1	0.4

我们以 MNIST 为例，图 3.1给出了 GRF 在 MNIST 数据集上的测试结果。图 3.2给出了 GRF 精度随训练集比例的曲线。图 3.3给出了 LLGC 在 MNIST 数据集上的测试结果。图 3.4分别给出了 LLGC 精度随超参 σ 和 λ 的测试结果。

图 3.5和图 3.6与上相同，分别给出 Vowel 和 Sonar 的结果。

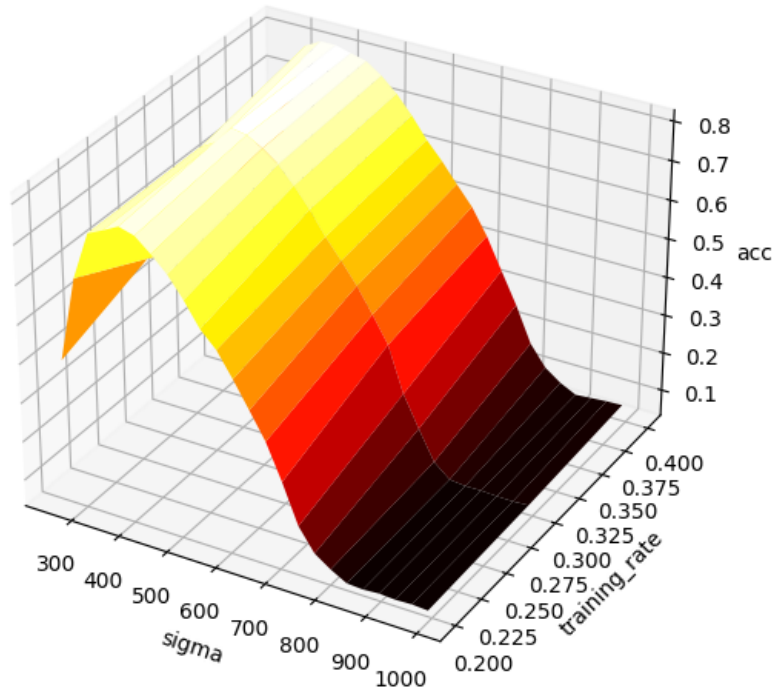


图 3.1 GRF 在 MNIST 的测试结果

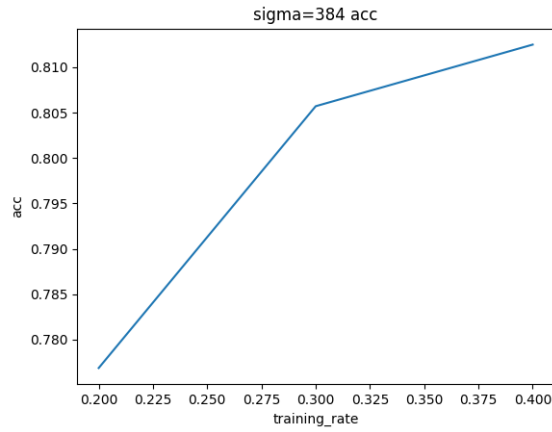


图 3.2 GRF 精度随训练集比例的实验结果

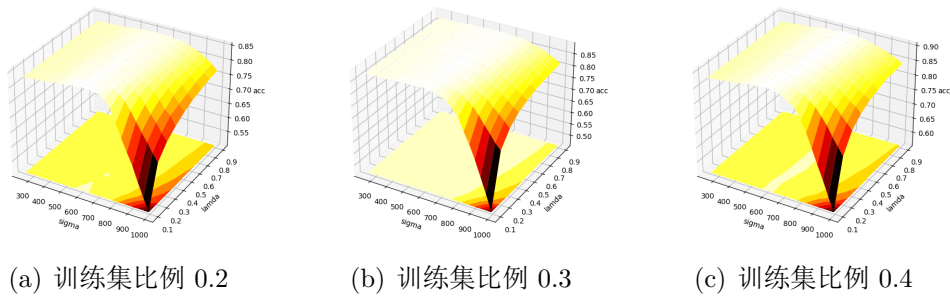


图 3.3 LLGC 在 MNIST 的测试结果

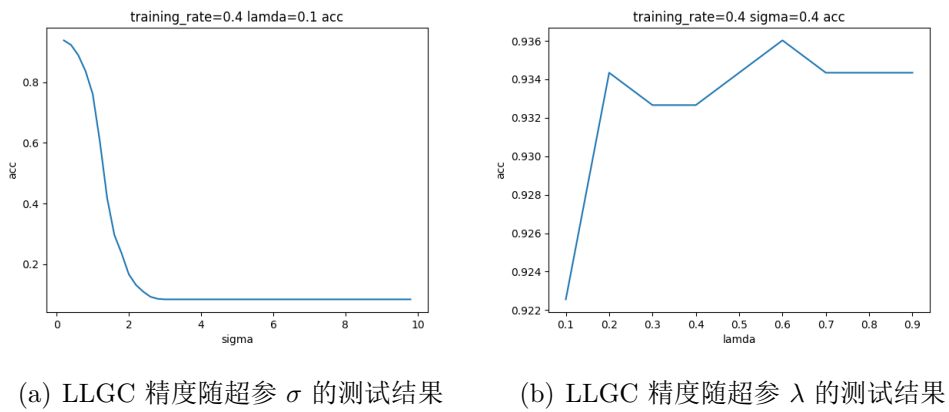
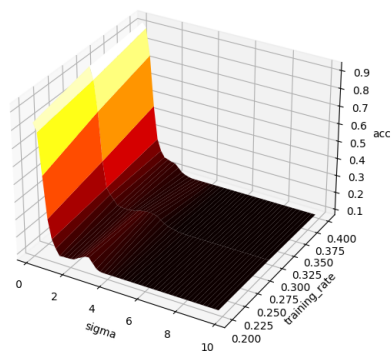


图 3.4 LLGC 精度随超参 σ 和 λ 的测试结果

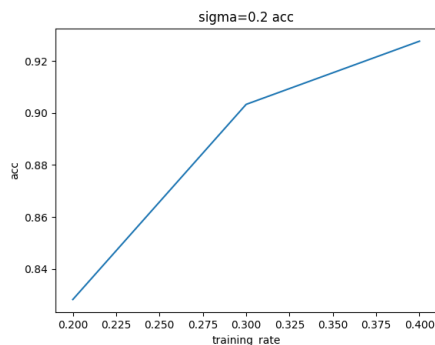
3.3 分析

从表 3和上节图中可以分析得到以下结论：

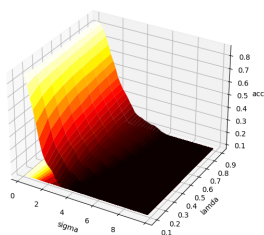
1. LLGC 方法的确优于 GRF，在 MNIST 和 Vowel 数据集上大幅领先，在



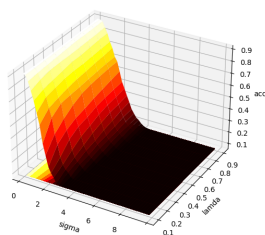
(a) GRF 在 Vowel 的测试结果



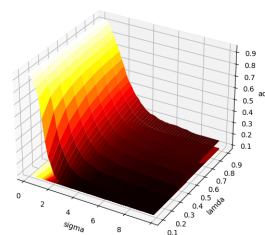
(b) GRF 精度随训练集比例的实验结果



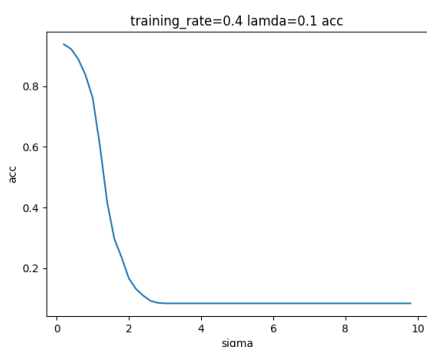
(c) 训练集比例 0.2



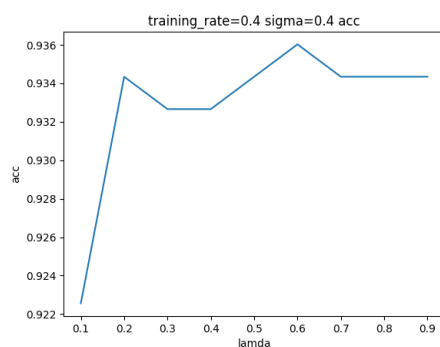
(d) 训练集比例 0.3



(e) 训练集比例 0.4



(f) LLGC 精度随超参 σ 的测试结果



(g) LLGC 精度随超参 λ 的测试结果

图 3.5 GRF 及 LLGC 在 Vowel 的测试结果

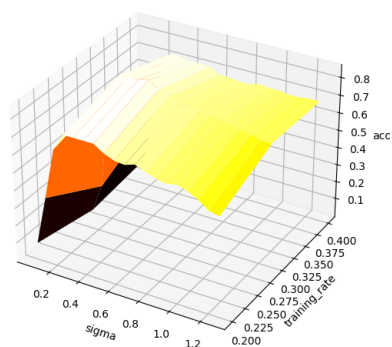
Sonar 数据集上持平。

2. 半监督训练中，训练集所占比例越大，最终测试集结果越好 (Sonar 数据集除外)。

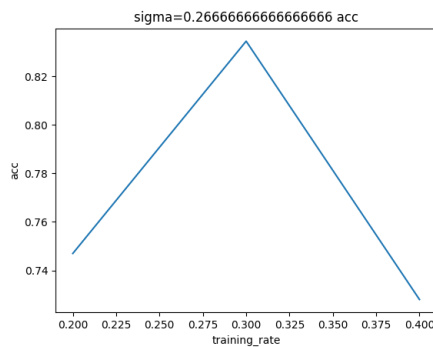
3. 超参数 σ 的影响

对于 GRF 和 LLGC，我们发现三个数据集上都经历了效果提升，到达最佳，效果下降的过程，即 σ 不宜取值过大，也不宜过小，需要根据数据集上样本特征向量中每一维元素的具体值并结合循环调参确定。

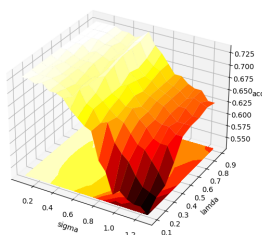
4. 超参数 λ 的影响



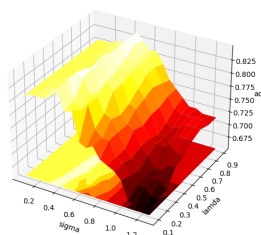
(a) GRF 在 Nosar 的测试结果



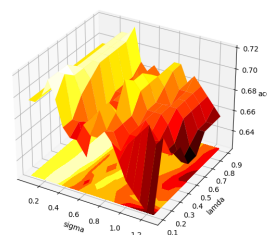
(b) GRF 精度随训练集比例的实验结果



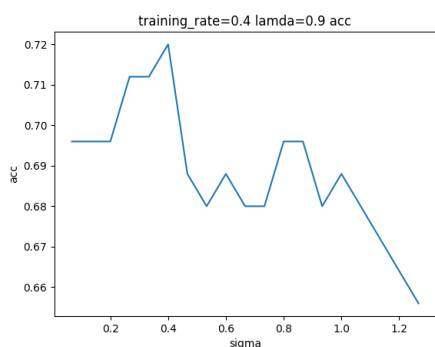
(c) 训练集比例 0.2



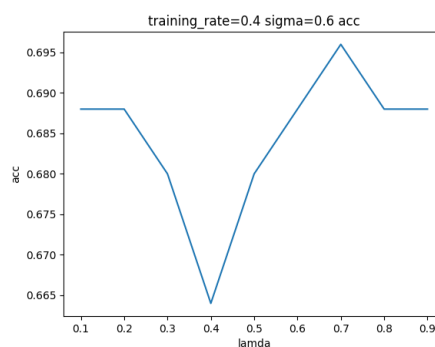
(d) 训练集比例 0.3



(e) 训练集比例 0.4



(f) LLGC 精度随超参 σ 的测试结果



(g) LLGC 精度随超参 λ 的测试结果

图 3.6 GRF 及 LLGC 在 Nosar 的测试结果

对于 GRF 和 LLGC, 我们发现三个数据集上的 λ 大部分情况下都是越小效果越好, 这里 λ 代表能量函数式 7 中的第二项系数, 本来这一系数应该越大越好, 代表在已知标签样本上新标签和原始真实标签越接近, 但是由于这里把未标签样本也考虑了进去, 因此这一系数越小越好。

第 4 节 小结

本次大作业旨在对两种半监督方法: GRF 和 LLGC 进行分析。首先将 3 组数据集 MNIST、Vowel 和 Sonar 构建为图连接数据形式, 再根据半监督训练范

式，通过实验得到了分类精度，分析了超参数对结果的影响，最终实验结果符合预期，较好得完成了本次大作业。

第 5 节 参考文献

- [1] Dengyong Zhou, Olivier Bousquet, Thomas Navin Lal, Jason Weston, and Bernhard Sch Olkopf. Learning with local and global consistency. *Advances in neural information processing systems*, 16(3), 2003.
- [2] Xiaojin Zhu, Zoubin Ghahramani, and John D. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *Machine Learning, Proceedings of the Twentieth International Conference (ICML 2003), August 21-24, 2003, Washington, DC, USA*, 2003.

第 6 节 代码

代码已经上传至本人 GitHub 仓库，详见
https://github.com/huaizepeng2020/PRML_HW3.git。

其中，`data1.py`、`data2.py` 和 `data3.py` 分别代表在 MNIST、Vowel 和 Sonar 上测试 GRF 和 LLGC。`plot_data.py` 对所有实验结果画图。