

基于梯度下降的不可微损失函数优化算法

薛艳锋, 刘继华, 张 翔, 薛志文

(吕梁学院计算机科学与技术系, 山西 吕梁 033000)

✉ 644126935@qq.com; 157598066@qq.com; 471623290@qq.com; 421070155@qq.com



摘 要: 由于谱半径与矩阵的映射关系无法用一个可微函数显式表示, 所以无法直接利用梯度下降算法进行计算。针对这一问题, 提出一种基于梯度下降的不可微损失函数优化算法。首先, 利用矩阵的 F 范数替代谱半径构建损失函数。其次, 基于谱半径小于等于 F 范数的事实, 构建初始化参数矩阵进而计算目标矩阵。最后, 如果目标矩阵的谱半径小于阈值, 则参数矩阵停止更新。实验结果表明, 与随机连边、度小优先连边及度大优先连边相比, 基于梯度下降的连边数量更多。

关键词: 感染强度; 传染阈值; 梯度下降; 谱半径; F 范数

中图分类号: TP181 **文献标识码:** A

Optimization Algorithm of Non-differentiable Loss Function Based on Gradient Descent

XUE Yanfeng, LIU Jihua, ZHANG Xiang, XUE Zhiwen

(Department of Computer Science and Technology, Lvliang University, Lvliang 033000, China)

✉ 644126935@qq.com; 157598066@qq.com; 471623290@qq.com; 421070155@qq.com

Abstract: The mapping relationship between spectral radius and matrix cannot be expressed by a differentiable function, so it cannot be directly calculated by gradient descent algorithm. To solve this problem, this paper proposes a non-differentiable loss function optimization algorithm based on gradient descent. Firstly, the loss function is constructed by replacing the spectral radius with the F -norm of the matrix. Secondly, since the spectral radius is less than or equal to the F -norm, the initialization parameter matrix is constructed to calculate the target matrix. Finally, if the spectral radius of the target matrix is less than the threshold, the parameter matrix stops updating. The experimental results show that the number of connected edges based on gradient descent is more than that based on random connected edges, low degree first connected edges and high degree first connected edges.

Keywords: infection intensity; threshold of infection; gradient descent; spectral radius; F -norm

1 引言(Introduction)

关于流行病动力学的数学模型是了解和干预流行病在现实世界传播的必要工具^[1]。比如, SIS (Susceptible-Infected-Susceptible) 模型可以刻画个体被感染、恢复再被感染、恢复的周而复始的流行病传播过程^[2]。接触网络(节点表示个体, 连边表示个体之间的接触)可以更真实地刻画现实世界流行病的传播过程^[3]。在 SIS 模型中, 如果感染强度大于传染阈值, 则传染病会流行, 反之, 则不会流行; 其中, 传染阈值为接

触网络谱半径的倒数^[4]。然而现实世界中, 不可能存在一个孤立的群体(用接触网络 G 表示)。如果接触网络 G_1 与另一个接触网络 G_2 相互连接, 则彼此的传染阈值都会减小^[4], 进而可能导致相互连接之前都不会暴发的流行病在连接之后暴发。所以, 在连接之后仍然不会暴发流行病的条件下, 求取最大连边数很有必要。该过程可通过梯度下降算法计算, 但是矩阵到谱半径的映射无法用可微函数表示, 所以无法直接应用。本文利用目标矩阵的 F 范数(该 F 范数大于谱半径且是矩阵的

凸函数)构建损失函数,并初始化连接矩阵的参数满足特定分布,进而通过梯度下降算法成功找到满足约束条件的最优解。

2 预备知识与问题定义(Preliminary knowledge and definition of problem)

2.1 基于接触网络的 SIS 模型

在基于单个接触网络 G (用邻接矩阵 A 表示,元素为 1 表示接触,为 0 表示未接触)的 SIS 模型中,个体分为感染者(I)和易感者(S),如图 1 所示,恢复率 δ 为感染者恢复为易感者的概率(如感染者 2、4 所示),感染率 β 为感染者通过接触(用连边表示)成功感染邻居易感者的概率(如感染者 7 成功感染邻居易感者 4)。其中,感染强度 τ 和传染阈值 ϵ 定义如下:

$$\tau = \beta / \delta \quad (1)$$

$$\epsilon = 1 / \rho(A) \quad (2)$$

其中, $\rho(A)$ 为接触网络 G 的谱半径。

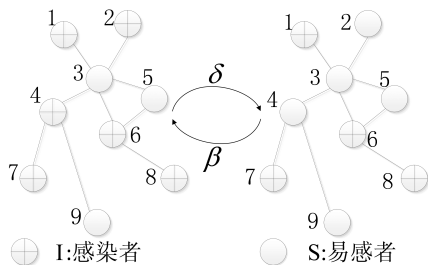


图 1 基于接触网络的 SIS 模型

Fig. 1 SIS models based on the contact networks

如图 2 所示,当接触网络 G_1 (黑色表示)与另一个接触网络 G_2 (灰色表示)相互连接时,分别用 A_{11} 和 A_{22} 表示接触网络 G_1 和 G_2 的邻接矩阵, N_1 和 N_2 表示节点个数, δ_1 和 δ_2 表示恢复率, A_{12} 表示 G_1 和 G_2 之间的个体接触(黑色虚线表示)。 β_{ij} 表示 G_j 对 G_i 的感染率($i, j \in \{1, 2\}$)。此外,有以下几点需要补充。

(1) 本文所有的网络为无向网络,则有 $A_{11} = A_{11}^T$, $A_{22} = A_{22}^T$, $A_{12} = A_{21}^T$, 其中矩阵元素 $(A_{12})_{ij}$ 表示 G_1 的第 i ($i \in \{1, 2, \dots, N_1\}$) 个节点与 G_2 的第 j ($j \in \{1, 2, \dots, N_2\}$) 个节点接触。

(2) 由于感染的过程同时涉及感染者和易感者,所以感染率 β 可分解为感染者呼出病毒的速率 μ 和易感者吸入病毒的概率 ω 的乘积,则有

$$\beta_{11} = \mu_1 \omega_1 \quad (3)$$

$$\beta_{22} = \mu_2 \omega_2 \quad (4)$$

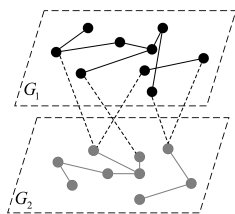


图 2 相互连接的接触网络

Fig. 2 An interconnected contact network

2.2 定义

无标度网络^[5]: 通过生长与偏好连接生成的网络,具体

过程为每次输入 m 条边与已经存在的 m 个节点相连(生长),偏好连接体现在优先连接度大的节点。由于无标度网络更能代表真实网络^[6],所以本文设定 G_1 和 G_2 都为无标度网络,通过 networkx 库^[7]生成。

谱半径: 矩阵 $A \in \mathbb{C}^{N \times N}$ 的谱半径表达式为 $\rho(A) = \max(|\lambda_1|, \dots, |\lambda_N|)$, 即矩阵特征值模的最大值,又因为本文假定所有网络为无向网络,即邻接矩阵为对称矩阵,所以谱半径可简化为邻接矩阵的最大特征值。

矩阵的 F 范数 $\|A\|_F$: 把矩阵中每个元素的平方求和再开根号,即

$$\|A\|_F = \sqrt{\sum_i \sum_j (A_{ij})^2} \quad (5)$$

由于连接矩阵 A_{12} 的元素取值只能为 0 或 1, 所以 G_1 和 G_2 的接触连边数可表示为 $\|A_{12}\|_F^2$, 即 F 范数的平方。

矩阵不等式 $W \geq A$: 对于任意 $N_1 \times N_2$ 的同型矩阵 W 和 A , 对应元素都使不等式 $W_{ij} \geq A_{ij}$ ($i \in \{1, 2, \dots, N_1\}$, $j \in \{1, 2, \dots, N_2\}$) 成立, 则有矩阵不等式 $W \geq A$ 成立; 而对于非负同型矩阵 $W \geq A$, 则有 $\|W\|_F^2 \geq \|A\|_F^2$ 。

2.3 问题

G_1 和 G_2 相互连接之前, 感染强度和传染阈值分别如下:

$$\tau_{11} = \beta_{11} / \delta_1 \quad (6)$$

$$\tau_{22} = \beta_{22} / \delta_2 \quad (7)$$

$$\epsilon_{11} = 1 / \rho(A_{11}) \quad (8)$$

$$\epsilon_{22} = 1 / \rho(A_{22}) \quad (9)$$

其中, 当 $\tau_{11} < \epsilon_{11}$ 且 $\tau_{22} < \epsilon_{22}$ 时, G_1 和 G_2 都不会暴发流行病。文献[4]指出: G_1 连接 G_2 之后, G_1 的传染阈值下降:

$$\epsilon_{11,c} = 1 / \rho(H) \quad (10)$$

其中,

$$H = A_{11} + \alpha^2 \tau_{22} A_{12} (I - \tau_{22} A_{22})^{-1} A_{12}^T \quad (11)$$

本文研究的问题为公式(11)的逆问题,即在相同的前提下条件下($\tau_{11} < \epsilon_{11}$ 且 $\tau_{22} < \epsilon_{22}$), 计算满足 G_1 连接 G_2 之后, 仍然不会暴发流行病($\omega_{11} < \epsilon_{11,c}$)的最大接触边数 $\|A_{12}\|_F^2$, 该问题可形式化地表示为满足约束条件的最优化问题:

$$\|A_{12}^*\|_F^2 = \arg \max_{\|A_{12}\|_F^2} \min_{\epsilon_{11,c}} \quad (12)$$

$$\text{s. t. } \epsilon_{11,c} > \tau_{11} \quad (13)$$

依据公式(10)可转化为如下最优化问题:

$$\|A_{12}^*\|_F^2 = \arg \max_{\|A_{12}\|_F^2} \rho(H) \quad (14)$$

$$\text{s. t. } \rho(H) < 1 / \tau_{11} \quad (15)$$

3 算法(Algorithm)

深度学习的函数拟合过程都是利用损失函数通过梯度下降算法调整参数矩阵,从而在参数空间找到一组最优参数的过程^[8-11]。同理,本文也是利用损失函数在两个接触网络 G_1 和 G_2 的接触连边 A_{12} 空间中寻找一组最多连边的组合。

设矩阵 H 的谱半径和 F 范数与 G_1 感染强度的倒数差值分别如下:

$$d_{\rho(H)} = \rho(H) - 1 / \tau_{11} \quad (16)$$

$$d_{\|H\|_F} = \|H\|_F - 1 / \tau_{11} \quad (17)$$

由于矩阵到谱半径的映射无法用可微函数计算,所以本文采用公式(17)利用 F 范数替代公式(16)构建损失函数,具体流程如下。

(1)初始化矩阵 $\mathbf{W}_{N_1 \times N_2}$ 的元素满足区间为 $[1, 2]$ 的均匀分布, 替换矩阵 $\mathbf{A}_{12}$ 代入公式(11)可得矩阵 $\mathbf{H}_W$ 。

(2)设置矩阵 $\mathbf{A}_{12} = \mathbf{W} > 0.5$, 即

$$(\mathbf{A}_{12})_{ij} = \begin{cases} 1, & \mathbf{W}_{ij} > 0.5 \\ 0, & \text{其他} \end{cases} \quad (18)$$

其中, $i \in \{1, 2, \dots, N_1\}$, $j \in \{1, 2, \dots, N_2\}$ 。代入公式(11)可得矩阵 $\mathbf{H}$ 。此时, 由于矩阵 $\mathbf{W} > \mathbf{A}_{12}$ (矩阵 $\mathbf{A}_{12}$ 表示元素全为1的矩阵), 所以可得矩阵不等式 $\mathbf{H} < \mathbf{H}_W$ 成立, 进而 $\rho(\mathbf{H}) < \|\mathbf{H}_W\|_F$ 。

(3)设计初始化参数矩阵 $\mathbf{W}_{N_1 \times N_2}$ 的指导原则是使 G_1 的感染强度倒数小于矩阵 $\mathbf{H}$ 的谱半径 $\rho(\mathbf{H})$, 即有

$$1/\tau_{11} < \rho(\mathbf{H}) \quad (19)$$

进而, 不等式

$$0 < 1/\tau_{11} < \rho(\mathbf{H}) < \|\mathbf{H}_W\|_F \quad (20)$$

成立。

(4)利用公式(17)构建损失函数 $L = d_{\rho(\mathbf{H})}^2 \|\mathbf{H}_W\|_F$ 。

(5)通过梯度下降算法更新参数矩阵 $\mathbf{W}$ 。利用公式(20)表示的不等式关系, 当损失值不断减小时, 公式(16)表示的 $d_{\rho(\mathbf{H})}$ 不断接近于0。当首次 $d_{\rho(\mathbf{H})} < 0$ 时, 也就是公式(15)表示的约束条件初次得到满足, 此时公式(14)表示的最优解 $\|\mathbf{A}_{12}^*\|_F^2$ 成功获得, 即在 G_1 连接 G_2 之后, 仍然不会暴发流行病前提下, 对应的最大连边数为 $\|\mathbf{A}_{12}^*\|_F^2$, 同时对应的连边方式 $\mathbf{A}_{12}^*$ 也唯一确定。

4 实验结果及分析(Experimental results and analysis)

4.1 系数设置

本文选择的系数设置如表1所示, N 表示节点数, m 表示无标度网络每次生长的边数, τ_{22} 表示 G_2 的感染强度, “—”表示占位符, 不代表任何实际含义。

取两种极端情况如下。

(1)当 G_1 与 G_2 相互连接之前(即 $\|\mathbf{A}_{12}\|_F^2 = 0$, 连边矩阵 $\mathbf{A}_{12}$ 的元素都为0), 矩阵 $\mathbf{H}$ 与 G_1 的邻接矩阵 $\mathbf{A}_{11}$ 相等, 此时, $\mathbf{H}$ 矩阵的谱半径最小, 即

$$\min \rho(\mathbf{H}) = \rho(G_1) = 7.53 \quad (21)$$

这种情况下, G_1 的传染阈值最大, 即

$$\max \epsilon_{11,c} = 1/7.53 \quad (22)$$

此时, 表示 G_1 与 G_2 没有任何连接, 即针对矩阵 $\mathbf{H}$ 的研究等价于研究完全孤立的群体 G_1 , 所以这种情况不属于本文研究的范围。

(2)当 G_1 与 G_2 相互全连接之后(即 $\|\mathbf{A}_{12}^*\|_F^2 = N_1 \times N_2 = 200 \times 120 = 24\,000$, 连边矩阵 $\mathbf{A}_{12}$ 的元素都为1), 矩阵 $\mathbf{H}$ 的谱半径最大, 即

$$\max \rho(\mathbf{H}) = \rho(G_1) = 105.80 \quad (23)$$

这种情况下, G_1 的传染阈值下降为最小值:

$$\min \epsilon_{11,c} = 1/105.80 \quad (24)$$

如果在这种情况下仍然满足前提条件 $\tau_{11} < \min \epsilon_{11,c}$, 即仍然不会暴发传染病, 那么可以把 G_1 与 G_2 结合起来统一看作孤立群体 G , 所以这种情况也不属于本文研究的范围。

综合上述两种情况, 本文研究的 G_1 感染强度的倒数 $1/\tau_{11}$ 范围:

$$1/\tau_{11} \in (1/\max \epsilon_{11,c}, 1/\min \epsilon_{11,c}) = (7.53, 105.80) \quad (25)$$

基于此, 本文选择 G_1 感染强度的倒数 $1/\tau_{11}$ 分别为 20、40、60、80。

表1 系数设置

Tab.1 Coefficients settings

G_1		G_2	
示数	取值	示数	取值
N_1	200	N_2	120
m_1	2	m_2	1
—	—	τ_{22}	$1/[2 \times (N_2 - 1)]$

4.2 实验方案

实验方案分为四种: 方案一为梯度下降、方案二为随机连边、方案三为度大优先(连边)、方案四为度小优先(连边)。其中, 方案一中, 学习率设置为 0.000 1; 方案二中, 每条连边都等概率选择 G_1 和 G_2 中的节点; 方案三中, 按照无标度网络的方式选择节点; 方案四中, 每个节点被选择的概率为 0.95^{k_i} , 其中 k_i 表示节点 i 的度。按照方案二、方案三及方案四三种方案增加连边, 直到 $\rho(\mathbf{H}) > 1/\tau_{11}$, 即 G_1 暴发流行病为止, 则最近上一次的连边数量即为 $\|\mathbf{A}_{12}^*\|_F^2$ 。

4.3 实验结果

实验目的是在 $\tau_{11} < \epsilon_{11}$ 且 $\tau_{22} < \epsilon_{22}$ 的前提下, 计算满足公式(15)的最大接触边数 $\|\mathbf{A}_{12}^*\|_F^2$, 该值越大, 说明对应算法越有效; 表示形式为重复 10 次实验结果的均值 ± 标准差, 均值越大, 说明算法越有效, 标准差越小, 说明算法越稳定。

如表2所示, 在四种 G_1 感染强度的倒数 $1/\tau_{11}$ 的不同取值情况下, 梯度下降的方法计算得到满足条件的连边数最大及标准差最小。同时, 当 G_1 的感染强度越大(越容易暴发流行病, 即 $1/\tau_{11}$ 越小), 本文所提算法取得的效果越明显。

表2 实验结果对比

Tab.2 Comparison of experimental results

$1/\tau_{11}$	随机连边	度小优先	度大优先	梯度下降
20	9 187.8 ± 12.4	9 259.4 ± 10.3	7 377.6 ± 121.3	9 402.7 ± 3.0
40	14 145.4 ± 8.0	14 198.2 ± 3.5	12 500.2 ± 126.2	1 4227.4 ± 2.9
60	17 734.2 ± 4.3	17 765.3 ± 4.4	16 693.8 ± 71.3	17 774.6 ± 3.4
80	20 706.4 ± 3.4	20 721.0 ± 1.2	20 222.8 ± 57.2	20 727.0 ± 1.2

4.4 理论分析

本文研究的核心思想是初始化参数矩阵 $\mathbf{W}_{N_1 \times N_2}$, 在此基础上, 通过添加判断条件获得连边矩阵 $\mathbf{A}_{12}$ 以使公式(20)成立, 最后利用该不等式构建损失函数。随着训练轮数的增加, 参数矩阵 $\mathbf{W}_{N_1 \times N_2}$ 的元素不断变化, 进而 $\|\mathbf{H}_W\|_F$ 和 $\rho(\mathbf{H})$ 逐步靠近 $1/\tau_{11}$ 。如图3所示, 在寻找到最大接触连边 $\|\mathbf{A}_{12}^*\|_F^2$ 之前, 公式(20)的不等式条件始终成立, 其中的画圈仅仅出于可视化的目的, 为了说明 $\rho(\mathbf{H})$ 值开始下降时, 公式(20)的不等式条件仍然成立。不可微损失函数可以通过梯度下降算法进行优化, 说明本文算法设计的合理性, 这是本文算法的创新点之一。此外, 在深度学习中, 参数更新停止的条件

大体为评价模型的负向指标(如损失值)在训练集上继续减小,而在验证集上停止减小;或者与之对应的正向指标(如准确度)在训练集上继续增大,而在验证集上停止增大。训练出来的模型如果在测试集上的效果不理想时,尽管可以利用网格搜索调整模型的超参数,但仍然是一个非常烦琐、耗时及反复的过程。本文所提算法在公式(20)表示的不等式前提下,程序一直运行,退出循环之时,也是优化结果成功获得之时。相比现今主流深度学习算法需要不断调参的反复探索过程,这种能获得最优计算结果的确定性无疑极大地缩短了算法运行的时间。

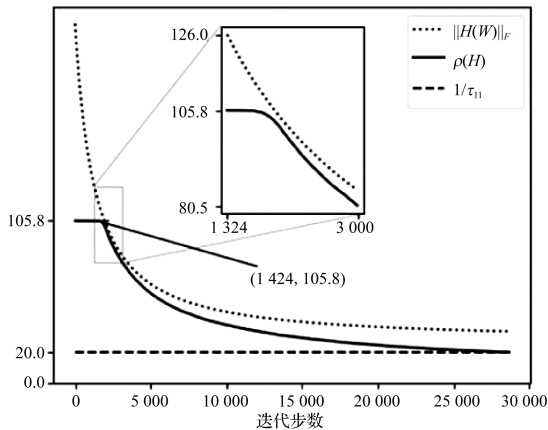


图3 当 $1/\tau_{11}=20$ 时, 谱半径和 F 范数与感染强度倒数变化示意图

Fig. 3 Schematic diagram of reciprocal changes of spectral radius and F norm with infection intensity when $1/\tau_{11}=20$

4.5 算法运行过程描述

从图3可以看出,代表 H 矩阵谱半径 $\rho(H)$ 的曲线在训练初期,谱半径的值处于最大值 105.8,此时对应的连边矩阵 A_{12} 的元素都为 1,即 G_1 与 G_2 中的每个个体之间都相互连接。随着训练的持续进行,谱半径 $\rho(H)$ 开始下降,即去边的过程开始(开始时的坐标通过元组显示),直到优化过程结束。

4.6 算法的扩展性分析

本算法的扩展性从三个方面进行分析。

(1)算法层面:如果 A_{12} 表示接触概率,则其他三种连边方式失效,本文算法只需要把公式(18)替换为 $(A_{12})_{ij} = \text{sigmoid}(W_{ij})$ 即可。

(2)硬件层面:可以充分利用图形处理器(Graphics Processing Unit, GPU)并行加速计算,对于规模较大的网络,可以明显提高运算效率。

(3)软件层面:各种深度学习的开源框架 TensorFlow^[12] 和 PyTorch^[13] 不仅支持 GPU 运算,而且提供了一系列完整的损失函数、初始化参数矩阵、自适应学习率算法等,这些都为本文所提算法的扩展提供了便利。

5 结论(Conclusion)

本文利用矩阵的 F 范数替代谱半径构建损失函数,成功地解决了两个接触网络的最多连边优化问题,并从实验结果

和理论分析证明了所提算法的有效性。而且,通过算法运行过程的描述,直观地展示了寻找最多连边的整个过程并能对关键环节给出解释。此外,从算法层面、硬件层面以及软件层面对所提算法的扩展性进行了分析,为解决不可微损失函数的优化问题提供了新的思路。

参考文献(References)

- [1] KATTIS A A, HOLIDAY A, STOICA A A, et al. Modeling epidemics on adaptively evolving networks: a data-mining perspective [J]. Virulence, 2015, 7(2):153-162.
- [2] 罗晓峰,李星.异质网络上 SIS 传染病模型两种逼近方法的比较[J].重庆理工大学学报(自然科学),2021,35(9):214-219.
- [3] 方鼎.基于自适应网络的传染病 SIS 模型动力学分析[D].重庆:西南大学,2020.
- [4] SAHNEH F D, SCOGGIO C, CHOWDHURY F N. Effect of coupling on the epidemic threshold in interconnected complex networks: A spectral analysis [C]// IEEE. 2013 American control conference, Washington, DC, USA: IEEE, 2013:2307-2312.
- [5] BARABASI A. Network Science[M]. Cambridge: Cambridge University Press, 2016:86-87.
- [6] FAN C, ZENG L, SUN Y. et al. Finding key players in complex networks through deep reinforcement learning[J]. Nat Mach Intell, 2020, 2:317-324.
- [7] HAGBERG A, SWART P, CHULT D. Exploring network structure, dynamics, and function using NetworkX[R]. Los Alamos National Lab. (LANL), Los Alamos, NM (United States), 2008.
- [8] 冯慧敏,金崑.基于卷积神经网络的雄性海南长臂猿声纹识别[J].林业科学,2023,59(1):119-127.
- [9] 张宇,温光照,米思娅,等.基于深度学习的二维人体姿态估计综述[J].软件学报,2022,33(11):4173-4191.
- [10] WU Z, PAN S, CHEN F, et al. A Comprehensive Survey on Graph Neural Networks [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1):4-24.
- [11] 史彩娟,任弼娟,王子雯,等.基于深度学习的伪装目标检测综述[J].计算机科学与探索,2022,16(12):2734-2751.
- [12] 陈锐,孙羽菲,程大果,等.TensorFlow中OpenCL核函数的实现与优化[J].计算机学报,2022,45(11):2456-2474.
- [13] 陈志华,杜垣均,温阳,等.基于注意力机制的密集残差融合与空间局部滤波低光照去雾算法[J].计算机辅助设计与图形学学报,2022,34(12):1842-1849.

作者简介:

薛艳锋(1984-),男,博士生,讲师.研究领域:图神经网络,传染病建模.

刘继华(1975-),女,硕士,教授.研究领域:软件测试,大数据智能分析.

张翔(2000-),男,本科生.研究领域:数据采集,大数据可视化.

薛志文(1982-),男,硕士,副教授.研究领域:计算机视觉,超分辨率重建.