

# 智能优化算法及其应用

## 文献阅读：模拟退火算法在深度学习中的应用

自 66 夏卓凡 2016011496

2019 年 5 月 5 日

### 1 引言

深度学习在近年来取得了举世瞩目的成果，不仅在学术研究上，而且在工程上的各个领域打造出了一些令人满意的产品。从 2012 年以来的这一波深度学习浪潮来看，现在取得良好效果的模型的优化方法都是使用基于梯度的优化器的反向传播算法，除此之外不借助梯度的算法十分少见。但也有将智能优化算法应用于深度学习之中的尝试，例如 Rere 等人 [1] 将模拟退火算法应用于 CNN 的优化实践。我将在这篇读书报告中简要阐述这篇文章的方法以及我个人的理解。

### 2 CNN 网络与模拟退火算法

卷积神经网络（CNN）作为图像领域的深度学习模型中不可或缺的一部分，已经作为一种层作为当前常见模型的组件之一。作为传统神经网络全连接层的补充，卷积层具有更加稀疏的连接，空间共享权重，易于捕捉图像中的局部模式等特征，在处理有关图像的问题上更加具有优势。

但是，不仅是卷积神经网络，所有深层神经网络的深度学习模型都存在着难以训练的问题。由于深度学习模型的参数量极大，使用二阶以上导数的优化算法，例如 Newton 迭代法等，都存在由于过高的时间复杂度和空间复杂度而难以实用的问题。常见的使用一阶导数的梯度优化器，如 SGD、Momentum、RMSProp、Adam 等，则存在对初始化极为敏感，不易收敛等问题。

智能优化算法分为基于单解的和基于种群的，基于种群的包括遗传算法 GA、差分进化算法 DE、粒子群算法 PSO 等，基于单解的算法包括模拟退火 SA、禁忌搜索 TS、噪声方法 NM 等。这篇文章使用了模拟退火算法代替了常见的梯度优化器对一个构建在 MNIST 训练集上的模型进行优化。

模拟退火算法最初被 Kirkpatrick 等人提出，在优化神经网络这一问题的具体的实现步骤为：

- (1) 产生初始化权重：按某种初始化方式初始化权重  $w_0$ ，并计算目标函数值  $f(x; w_0)$ ；
- (2) 设定初始温度：设定超参数——初温  $T$ ；
- (3) 开始迭代：从当前解  $w_t$  开始，选择一种扰动方式，构造新解  $w = w_t + \Delta w$ ；
- (4) 评估新解：如果  $f(x; w) < f(x; w_t)$ ，则接受解，令  $w_{t+1} = w$ ，否则按概率  $\exp(-\Delta f(x)/kT)$  接受，其中  $\Delta f(x) = f(x; w) - f(x; w_t)$ ， $k$  为 Boltzmann 常数，用于控制概率；若不接受，则  $w_{t+1} = w_t$ ；
- (5) 迭代寻求新解  $n$  次后，进行降温，开始新一轮迭代，转 (3) 降温方程  $T = cT$ ， $0 < c < 1$  为衰减因子；

(6) 当温度小于终止温度时，停止迭代。

在文章提出的方法中，用于优化的网络模型如下图所示。卷积层为 k5n6s1 和 k5n12s1，激活函数为 tanh 函数，下采样层使用步长为 2 的最大池化操作。

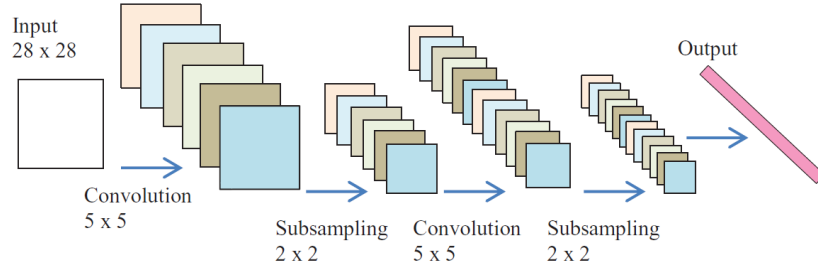


图 1: Rere 等人提出的网络架构 [1]

用于训练的目标函数非常简单，采用的是训练集 RMSE，即

$$f(o, y) = \frac{1}{2} \sqrt{\frac{1}{R} \sum_{i=1}^R (o - y)^2}$$

其中  $o$  为标签， $y$  为实际输出， $R$  为训练样本数量。当目标函数值小于一定值之后或者达到最大迭代步数之后算法终止。

### 3 结果与分析

文章给出了不同邻域大小（内循环迭代次数  $n$ ）的准确率和优化时间对比，以及和传统梯度方法的对比，如下所示。

Table 1. Performance comparison of cnn, cnnSA10, cnnSA20 and cnnSA50 in terms of accuracy

Methods	Number of epochs									
	1	2	3	4	5	6	7	8	9	10
cnn	88.87	92.25	93.90	94.81	95.44	96.18	96.59	96.92	97.22	97.27
cnnSA10	89.18	92.38	94.20	95.19	95.81	96.50	96.77	97.04	97.27	97.41
cnnSA20	90.50	92.77	94.68	95.45	96.66	96.87	97.08	97.26	97.30	97.61
cnnSA50	91.10	94.16	95.49	96.20	96.91	96.99	97.33	97.42	97.40	97.71

图 2: 准确率对比 [1]

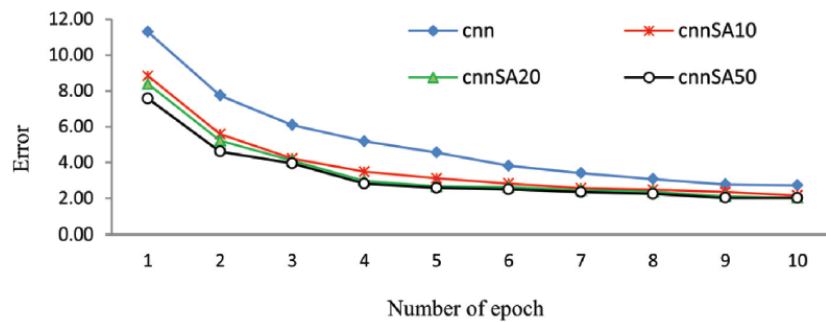


图 3: 错误率对比 [1]

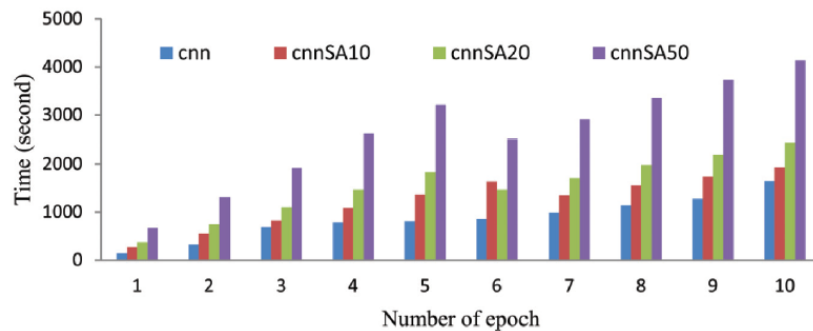


图 4: 优化时间对比 [1]

从以上的对比结果可以看出，模拟退火算法在固定的较小 epoch 时能得到比梯度下降算法更优的解，而且邻域越大，得到解的准确率越高；随着 epoch 增大，梯度下降算法趋向收敛，这时这些算法的表现十分接近；但是，模拟退火算法的优化时间要比梯度下降算法的时间显著增大，这或许是模拟退火等一类算法的 drawback。

## 4 个人的理解

本文提出了一种使用模拟退火算法优化神经网络的方法，可以说是一种有益的尝试。但本文在模型设计和实验分析上仍有一些问题，现对此进行分析。

首先模型设计上可以考虑使用 ReLU 作为激活函数，并使用 softmax-crossentropy 作为目标函数，这样的设置更适用于分类问题；其次在邻域中选取解的策略，文中为加上一个随机的扰动，但如果考虑 CNN 网络的特性，即卷积层提取特征，全连接层进行分类，那么在温度较高的时候，优先对卷积层权重加一些较大的扰动，而随着温度的下降，卷积层的扰动要迅速减小，在训练后期着重于全连接层的权重调整，这样的更新策略会比全都使用随机扰动更有针对性。

事实上，上面的结果也表明了使用梯度下降法优化神经网络，事实上最终的收敛解不会比用模拟退火算法差多少，得到的结果也是可以接受的；但是模拟退火算法在邻域较大的时候会比梯度下降算法耗时更长，而这可能是模拟退火算法等智能优化算法在工程上不常用的原因之一。

## 参考文献

- [1] LM Rasdi Rere, Mohamad Ivan Fanany, and Aniati Murni Arymurthy. Simulated annealing algorithm for deep learning. *Procedia Computer Science*, 72:137–144, 2015.