

denny的学习专栏

[博客园](#) [首页](#) [新随笔](#) [联系](#) [管理](#) [订阅](#) [XML](#)

随笔- 115 文章- 0 评论- 190

Caffe学习系列(7) : solver及其配置

solver算是caffe的核心的核心，它协调着整个模型的运作。caffe程序运行必带的一个参数就是solver配置文件。运行代码一般为

```
# caffe train --solver=*_solver.prototxt
```

在Deep Learning中，往往loss function是非凸的，没有解析解，我们需要通过优化方法来求解。solver的主要作用就是交替调用前向 (forward)算法和后向 (backward)算法来更新参数，从而最小化loss，实际上就是一种迭代的优化算法。

到目前的版本，caffe提供了六种优化算法来求解最优参数，在solver配置文件中，通过设置type类型来选择。

- Stochastic Gradient Descent (type: "SGD"),
- AdaDelta (type: "AdaDelta"),
- Adaptive Gradient (type: "AdaGrad"),
- Adam (type: "Adam"),
- Nesterov's Accelerated Gradient (type: "Nesterov") and
- RMSProp (type: "RMSProp")

具体的每种方法的介绍，请看本系列的下一篇文章, 本文着重介绍solver配置文件的编写。

Solver的流程：

1. 设计好需要优化的对象，以及用于学习的训练网络和用于评估的测试网络。（通过调用另外一个配置文件prototxt来进行）
2. 通过forward和backward迭代的进行优化来更新参数。
3. 定期的评价测试网络。（可设定多少次训练后，进行一次测试）
4. 在优化过程中显示模型和solver的状态

在每一次的迭代过程中，solver做了这几步工作：

- 1、调用forward算法来计算最终的输出值，以及对应的loss
- 2、调用backward算法来计算每层的梯度
- 3、根据选用的solver方法，利用梯度进行参数更新
- 4、记录并保存每次迭代的学习率、快照，以及对应的状态。

接下来，我们先来看一个实例：



```
net: "examples/mnist/lenet_train_test.prototxt"
test_iter: 100
test_interval: 500
base_lr: 0.01
momentum: 0.9
type: SGD
weight_decay: 0.0005
lr_policy: "inv"
gamma: 0.0001
power: 0.75
display: 100
max_iter: 20000
snapshot: 5000
snapshot_prefix: "examples/mnist/lenet"
solver_mode: CPU
```

昵称：[denny402](#)园龄：[5年10个月](#)粉丝：[74](#)关注：[2](#)[+加关注](#)

< 2016年5月 >						
日	一	二	三	四	五	六
24	25	26	27	28	29	30
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	1	2	3	4

搜索

<input type="text"/>	<input type="button" value="找找看"/>
<input type="text"/>	<input type="button" value="谷歌搜索"/>

常用链接

[我的随笔](#)[我的评论](#)[我的参与](#)[最新评论](#)[我的标签](#)[更多链接](#)

我的标签

[python\(26\)](#)[caffe\(25\)](#)[opencv3\(10\)](#)[matlab\(9\)](#)[mvc\(9\)](#)[MVC3\(8\)](#)[ajax\(7\)](#)[geos\(6\)](#)[opencv\(6\)](#)[ml\(5\)](#)[更多](#)

随笔分类

[caffe\(26\)](#)[GDAL\(2\)](#)[GEOS\(6\)](#)[matlab\(11\)](#)[opencv\(19\)](#)[Python\(25\)](#)

随笔档案

[2016年1月 \(33\)](#)[2015年12月 \(29\)](#)[2015年11月 \(10\)](#)[2015年7月 \(7\)](#)[2014年10月 \(4\)](#)

2016/5/11

Caffe学习系列(7)：solver及配置 - denny402 - 博客园



接下来，我们对每一行进行详细解读：

```
net: "examples/mnist/lenet_train_test.prototxt"
```

设置深度网络模型。每一个模型就是一个net，需要在一个专门的配置文件中对net进行配置，每个net由许多的layer所组成。每一个layer的具体配置方式可参考本系列文文章中的（2）-（5）。注意的是：文件的路径要从caffe的根目录开始，其它的所有配置都是这样。

也可用train_net和test_net来对训练模型和测试模型分别设定。例如：

```
train_net: "examples/hdf5_classification/logreg_auto_train.prototxt"
test_net: "examples/hdf5_classification/logreg_auto_test.prototxt"
```

接下来第二行：

```
test_iter: 100
```

这个要与test layer中的batch_size结合起来理解。mnist数据中测试样本总数为10000，一次性执行全部数据效率很低，因此我们将测试数据分成几个批次来执行，每个批次的数量就是batch_size。假设我们设置batch_size为100，则需要迭代100次才能将10000个数据全部执行完。因此test_iter设置为100。执行完一次全部数据，称之为一个epoch

```
test_interval: 500
```

测试间隔。也就是每训练500次，才进行一次测试。

```
base_lr: 0.01
lr_policy: "inv"
gamma: 0.0001
power: 0.75
```

这四行可以放在一起理解，用于学习率的设置。只要是梯度下降法来求解优化，都会有一个学习率，也叫步长。base_lr用于设置基础学习率，在迭代的过程中，可以对基础学习率进行调整。怎么样进行调整，就是调整的策略，由lr_policy来设置。

lr_policy可以设置为下面这些值，相应的学习率的计算为：

- fixed: 保持base_lr不变。
- step: 如果设置为step,则还需要设置一个stepsize, 返回 $\text{base_lr} * \gamma^{\lfloor \text{iter} / \text{stepsize} \rfloor}$, 其中iter表示当前的迭代次数
- exp: 返回 $\text{base_lr} * \gamma^{\text{iter}}$, iter为当前迭代次数
- inv: 如果设置为inv,还需要设置一个power, 返回 $\text{base_lr} * (1 + \gamma * \text{iter})^{-\text{power}}$
- multistep: 如果设置为multistep,则还需要设置一个stepvalue。这个参数和step很相似，step是均匀间隔变化，而multistep则是根据stepvalue值变化
- poly: 学习率进行多项式误差, 返回 $\text{base_lr} (1 - \text{iter} / \text{max_iter})^{\text{power}}$
- sigmoid: 学习率进行sigmoid衰减, 返回 $\text{base_lr} (1 / (1 + \exp(-\gamma * (\text{iter} - \text{stepsize}))))$

multistep示例：

```
base_lr: 0.01
momentum: 0.9
weight_decay: 0.0005
# The learning rate policy
lr_policy: "multistep"
gamma: 0.9
stepvalue: 5000
stepvalue: 7000
stepvalue: 8000
stepvalue: 9000
stepvalue: 9500
```



接下来的参数：

```
momentum : 0.9
```

上一次梯度更新的权重，具体可参看下一篇文章。

```
type: SGD
```

2014年7月 (4)

2013年10月 (3)

2013年8月 (5)

2013年7月 (7)

2013年6月 (6)

2011年4月 (4)

2010年6月 (3)

最新评论

1. Re:Caffe学习系列(23)：如何将别人训练好的model用到自己的数据上

您好，看到您的教程学到很多，我没用digs t，直接用的命令操作，但是由于电脑原因，我在做图片的Imdb的时候吧图片设置成125~125的，然后运行的时候就出现了错误错误内容是：Check failed.....

--weichang88688

2. Re:Caffe学习系列(4)：激活层（Activation Layers)及参数

给博主赞一个，对入门小白帮助真大！另外可以请问下你，为什么sigmoid层是另建一层，然后将自己输出，relu则本地操作不添加新的层，而后面的TanH,absolute value, power等都是.....

--MaiYatang

3. Re:Caffe学习系列(13)：数据可视化环境（python接口)配置

在哪个路径下Make Clear 呢？Caffe的编译会被清除么？

--TonyFaith

4. Re:Caffe学习系列(13)：数据可视化环境（python接口)配置

@TonyFaith清除以前的编译可以make clear，再重新编译就可以了。缺少python.h我不知道是什么原因...

--denny402

5. Re:Caffe学习系列(17)：模型各层数据和参数可视化

@weichang88688卷积层的输出数据就是net.blobs['conv1'].data[0]，用一个变量保存起来就可以了c1=net.blobs['conv1'].data[0]...

--denny402

阅读排行榜

1. SqlDataReader的关闭问题(9287)
2. 索引超出范围。必须为非负值并小于集合大小。(4655)
3. Caffe学习系列(1)：安装配置ubuntu14.04+cuda7.5+caffe+cuda(3166)
4. Caffe学习系列(12)：训练和测试自己的图片(2919)
5. Caffe学习系列(2)：数据层及参数(2492)

评论排行榜

1. Caffe学习系列(12)：训练和测试自己的图片(38)
2. SqlDataReader的关闭问题(22)
3. caffe windows 学习第一步：编译和安装（vs2012+win 64)(15)
4. Caffe学习系列(23)：如何将别人训练好的model用到自己的数据上(15)
5. Caffe学习系列(3)：视觉层（Vision Layers)及参数(11)

推荐排行榜

1. SqlDataReader的关闭问题(5)
2. Caffe学习系列(12)：训练和测试自己的图片(4)
3. Caffe学习系列(11)：图像数据转换成db（leveldb/lmdb)文件(2)

2016/5/11

Caffe学习系列(7)：solver及配置 - denny402 - 博客园

优化算法选择。这一行可以省掉，因为默认值就是SGD。总共有六种方法可选择，在本文的开头已介绍。

4. DWZ中刷新dialog的方案解决(2)
5. MVC3学习：利用mvc3+ajax结合MVCPa
ger实现分页(2)

```
weight_decay: 0.0005
```

权重衰减项，防止过拟合的一个参数。

```
display: 100
```

每训练100次，在屏幕上显示一次。如果设置为0，则不显示。

```
max_iter: 20000
```

最大迭代次数。这个数设置太小，会导致没有收敛，精确度很低。设置太大，会导致震荡，浪费时间。

```
snapshot: 5000  
snapshot_prefix: "examples/mnist/lenet"
```

快照。将训练出来的model和solver状态进行保存，snapshot用于设置训练多少次后进行保存，默认为0，不保存。snapshot_prefix设置保存路径。

还可以设置snapshot_diff，是否保存梯度值，默认为false,不保存。

也可以设置snapshot_format，保存的类型。有两种选择：HDF5 和BINARYPROTO，默认为BINARYPROTO

```
solver_mode: CPU
```

设置运行模式。默认为GPU,如果你没有GPU,则需要改成CPU,否则会出错。

注意：以上的所有参数都是可选参数，都有默认值。根据solver方法（type)的不同，还有一些其它的参数，在此不一一列举。

分类: [caffe](#)

标签: [caffe](#)

好文要顶

关注我

收藏该文



[denny402](#)

[关注 - 2](#)

[粉丝 - 74](#)

[+加关注](#)

1

推荐

0

反对

(请您对文章做出评价)

« 上一篇: [Caffe学习系列\(6\)：Blob,Layer and Net以及对应配置文件的编写](#)

» 下一篇: [Caffe学习系列\(8\)：solver优化方法](#)

posted @ 2015-12-24 19:26 [denny402](#) 阅读(1950) 评论(0) [编辑](#) [收藏](#)

[刷新评论](#) [刷新页面](#) [返回顶部](#)

注册用户登录后才能发表评论，请 [登录](#) 或 [注册](#)，[访问网站首页](#)。

最新IT新闻:

- Mac笔记本电脑出货量大幅跳水 相比去年跌40%
 - Slack开放第三方服务使用登入授权机制"Sign in with Slack"
 - 迪士尼财报低于预期，Disney Infinity 电玩产品线断头
 - 高晓松：我现在是wannabe企业家 未来要做真的企业家
 - 八成摄像头存安全隐患 家庭生活或被网上直播
- » [更多新闻...](#)

最新知识库文章:

- 架构漫谈（九）：理清技术、业务和架构的关系
 - 架构漫谈（八）：从架构的角度看如何写好代码
 - 架构漫谈（七）：不要空设架构师这个职位，给他实权
 - 架构漫谈（六）：软件架构到底是要解决什么问题？
 - 架构漫谈（五）：什么是软件
- » [更多知识库文章...](#)