

Physcalの大魔導書

传统弱校HFUT的蒟蒻，其幕后身份是，大魔导师(=￣ω￣=)。字体发虚右转MacType。

頁首 新章 微博 Github ModelZoo 文庫 祕境

Long-Short Memory Network(LSTM长短期记忆网络)

自剪枝神经网络

Simple RNN从理论上来看，具有全局记忆能力。因为T时刻，递归隐层一定记录着时序为1的状态

但由于Gradient Vanish问题，T时刻向前反向传播的Gradient在T-10时刻可能就衰减为0。

从Long-Term退化至Short-Term。

尽管ReLU能够在前馈网络中有效缓解Gradient Vanish，但RNN的深度过深，替换激活函数治标不治本。

$$\left| \prod_{j=p+1}^t \frac{\partial y_j}{\partial x_{t-1}} \right| \leq (\beta_W \cdot \beta_h)^{t-p} \quad \text{where } \beta = \text{UpperBound}$$

上式中显明的根源所在。由于W和h两个矩阵多次幂导致受数值影响敏感。简而言之就是深度过大。

大部分Long-Term情况下，不需要提供路径上完整的信息，但反向传播还是循规蹈矩地穿过这些冗深度。



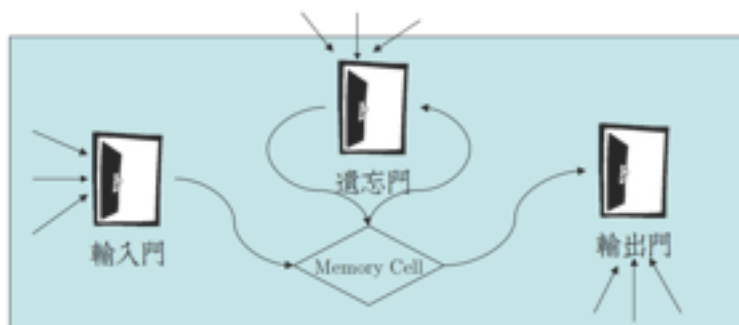
解决方案之一是，设置可自主学习的参数来屏蔽掉这些无用的信息，与“降维”相似，这种方法叫“降维”



神经网络的剪枝策略很简单，就是添加参数矩阵，经过一定周期的学习，选择性屏蔽掉输入，精简网络。

从结构上来看，类似“树套树”，就是“神经网络套神经网络”。

动态门结构



简单概括：

★LSTM将RNN的输入层，最前移入Memory Cell加以保护

★Input Gate, Forget Gate, Output Gate, 通过训练参数，将Gate或开(置1)或闭(置0)，保护Cell。

在时序展开图上则更加清晰：

公告



这是一个属于轻松描写魔导师平凡日常的故事，请不要过激期待。还有，请保持室内明亮离开电视3米以上再观看。(=￣ω￣=)

昵称：Physcal
园龄：1年7个月
粉丝：51
关注：19
+加关注

最新随笔

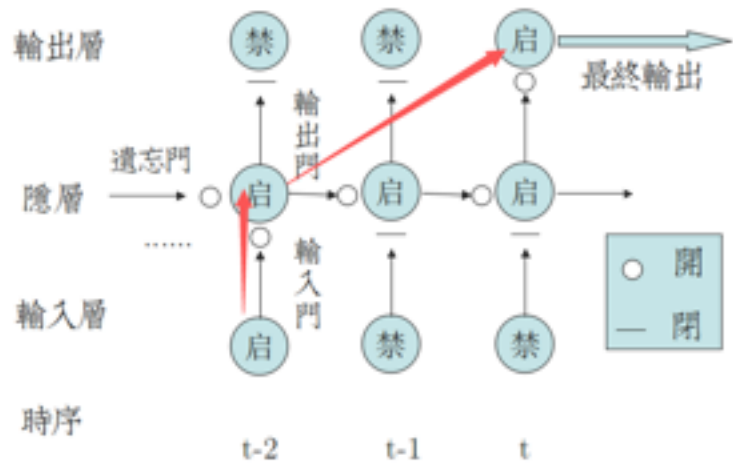
1. 从零开始山寨Caffe 抢笔...
2. 从零开始山寨Caffe 抢...
3. 从零开始山寨Caffe 抢...
4. 从零开始山寨Caffe 抢...
5. 从零开始山寨Caffe 抢...
6. 从零开始山寨Caffe 抢...
7. 从零开始山寨Caffe 抢...
8. 从零开始山寨Caffe 抢...
9. 从Bayesian角度浅析Bat...
10. 从零开始山寨Caffe 抢...

随笔分类 (165)

ACM(113)
Haskell(3)
Qt(1)
并行计算(3)
机器学习理论(29)
机器学习系统设计(12)
模式识别(4)

随笔档案 (161)

2016年3月 (8)
2016年2月 (5)
2015年11月 (1)
2015年10月 (2)
2015年9月 (2)
2015年8月 (7)
2015年7月 (1)
2015年6月 (8)
2015年5月 (19)
2015年4月 (3)
2015年3月 (7)
2015年2月 (10)
2014年11月 (17)
2014年10月 (71)



队友の魔導書

esegx
MaticsL
Pentium
战亿斯强

公式定义

原版LSTM最早在 [Hochreiter&Schmidhuber 97] 提出。

今天看到的LSTM是[Gers 2002]改良过的 *extended LSTM*。

extended LSTM扩展内容：

★Forget Gate，用于屏蔽t-1以及之前时序信息。

在时序展开图上，由左侧锁住以保护Cell。

★三态门控：

97年提出的Gate输入类似RNN，分为两态Weight矩阵：

- Wx——序列输入信息
- Wh——递归隐态输入信息

2002年补充了第三态：

- Wc——递归Cell态输入信息

将Cell的时序状态引入Gate，称为Peephole Weights。

唯一作用似乎是提升LSTM精度，Alex Graves的博士论文中这么说：

The peephole connections meanwhile improved the LSTM's ability to learn tasks that require precise timing and counting of the internal states.

具体实现的时候，为了增加计算效率，可以忽略：

Theano的Tutorial中这么说道：

The model we used in this tutorial is a variation of the standard LSTM model.

In this variant, the activation of a cell's output gate does not depend on the memory cell's state C_t .

This allows us to perform part of the computation more efficiently (see the implementation note, below, for details).

而CS224D Lecture8中压根就没提。

所以双态Gate可能是更为主流的LSTM变种。

2.1 前向传播

输入门：

$$i_t = \text{Sigmoid}(W_i x_t + U_i h_{t-1} + V_i C_{t-1}) \quad ①$$

遗忘门：

$$f_t = \text{Sigmoid}(W_f x_t + U_f h_{t-1} + V_f C_{t-1}) \quad ②$$

输出门：

$$O_t = \text{Sigmoid}(W_o x_t + U_o h_{t-1} + V_o C_t) \quad ③$$

原始CellRNN部分：

$$\tilde{C}_t = \text{Tanh}(W_c x_t + U_c h_{t-1}) \quad ④$$

门控Cell：

$$C_t = i_t \cdot \tilde{C}_t + f_t \cdot C_{t-1} \quad (\text{输入门} \cdot \text{遗忘门}) \quad ⑤$$
$$h_t = O_t \cdot \text{Tanh}(C_t) \quad \text{where } h_t = \text{FinalOutput} \quad (\text{输出门}) \quad ⑥$$

仔细观察①②③④，发现除了Peephole Weights引入的V项，这四个式子是一样的。

Theano中为了GPU能够一步并行计算，没有使用Peephole Weights，这样①②③这区就是一个基本并行模型：
以相同的代码，运算数据集在空间中的不同部分。

分类：机器学习理论

原文要图

关注数

收藏该文



Physical
关注 - 19
粉丝 - 51

+加关注

0

0

(请你对文章做出评价)

- 上一篇：[Recurrent Neural Network\(循环神经网络\)](#)
- 下一篇：[\[杂记\]机器学习：数学、概率与人工智能](#)

posted @ 2015-05-15 17:08 Physical 阅读(474) 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](#)。

- 【推荐】50万行VC++源码：大型城市工控、电力仿真CAD与GIS源码库
- 【推荐】跳出即时通讯云—芒果美食、Faces等亿级APP都在用
- 【推荐】怎样将“在线Excel”嵌入你的开发系统中？
- 【推荐】阿里云高可用云服务器+SSD云盘，让业务响应0延迟
- 【活动】当链家网的墙上高挂着大数据后台

野狗API
应用于各种实时通信场景

网络快，响应快，开发快

最新IT新闻

- 继百度云关闭后 百度相册也要关闭 将迁入百度云
- 2016年第一季度Android设备市场份额提升 苹果iOS萎缩
- 工信部：三大运营商提速降费还有大动作
- 陌陌李福海解：直播营收1560万美元 成最大收入来源
- 百合网澄清：与世纪佳缘合并交易还没有正式完成
- 更多新闻...

90%的开发者都在用 **极光推送**

—— 不只是稳定 ——

最新知识类文章

- 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代码
- 架构漫谈（七）：不要自设架构的这个职位，给谁实权
- 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
- 更多知识类文章...