

CNN

同时对卷积神经网络的算法结构和运用方式提出**针对性改进**，并将**研究和改进结果**运用于实际**高频股价数据**中，以预测.....

优化方向与思路？改进结果？

- 核结构
- 网络结构

1*1的卷积核有什么用，3*3的卷积核和一个1*3加一个3*1的有什么区别

1*1卷积可以改变上一层网络的**通道数目**。卷积核大于1*1，意味着**提特征需要邻域信息**。

- 若提取横向纹理，则横向邻域信息密度比纵向信息密度高。
- 核扁横着最科学。若提纵向纹理，同理，瘦高竖着最好。
- 若你想提取的纹理种类丰富，那横向邻域信息密度的期约等于纵向信息密度期望

所以对懒人来说，最优核的尺寸的期望是正方形。至于1*n和n*1，它们一般是搭配使用的，从而实现n*n卷积核的感受野，可以在减少参数的同时增加层数，在CNN的较高层中使用可以带来一定的优势。卷积核并非都是正方形的，还可以是矩形，比如3*5，在文本检测和车牌检测当中就有应用，这种设计主要针对文本行或者车牌的形状，能更好的学习特征。（其实方形矩形影响不大，网络的学习能力是非常强的。）

CNN是否适用于量化领域？适用于哪些细分？

- order-book建模
- k线形态（作为初步输入特征抽取，降维输入）
-

RNN

文本情绪分类，学习乐谱生成音乐

CNN和RNN异同？为什么要Normalization？Batch Norm和Layer Norm为什么分别用在CNN和RNN上？

相同

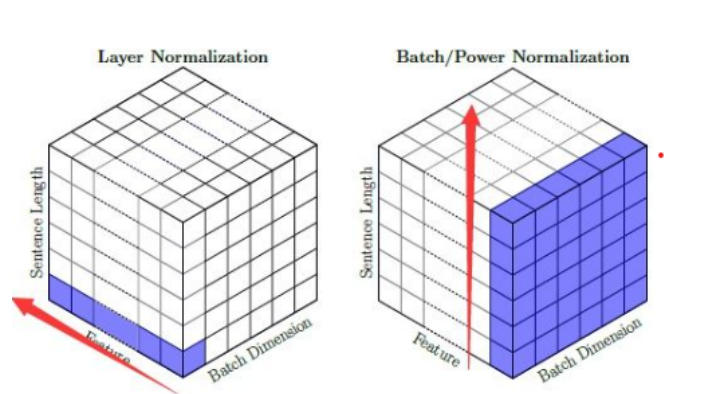
- 前向计算产生结果，反向计算模型更新
- 每层神经网络横向可以多个神经元共存,纵向可以有多层神经网络连接

不同

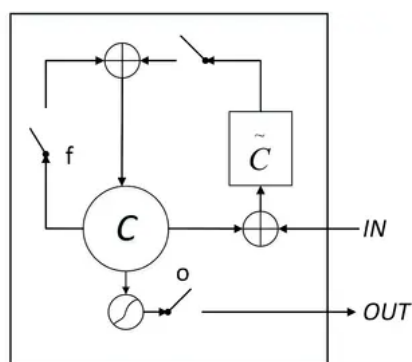
- CNN**空间扩展**，神经元与特征卷积；RNN**时间扩展**，神经元与多个时间输出计算
- RNN可以用于描述时间上连续状态的输出，有记忆功能，CNN用于静态输出

Norm

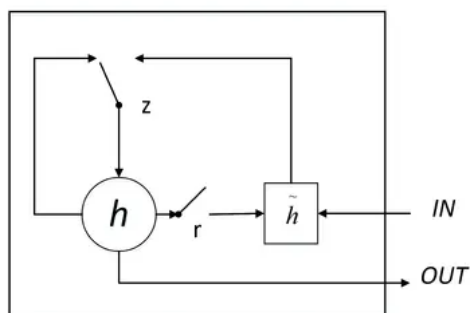
- Normalization的目的使得神经网络在计算过程中参数不落入**激活函数的饱和区**导致**训练性能下降**（参数投影变换为一个超球体）
- Batch Normalization 的处理对象是对**一批样本**， Layer Normalization 的处理对象是**单个样本**。
- Batch Normalization 是对这批样本的**同一维度特征**做归一化， Layer Normalization 是对这单个样本的**所有维度特征**做归一化
- BN会使得**批次间的关联**得以保留，而丢弃了**通道间的关联性**，在CNN主导的任务（类似CV）中保留批次间相关性更有意义，而在NLP领域，LN就更加合适。因为它抹杀了**不同样本间的大小关系**，但是保留了一个样本内**不同特征之间**的大小关系。对于NLP或者序列任务来说，一条样本的不同特征，其实就是时序上embedded feature上的时序变化，是十分紧密的关系应该保留。



LSTM和GRU的出现原因？异同？



(a) Long Short-Term memory



(b) Grated Recurrent Unit

- 均为了解决vanilla RNN的梯度爆炸和消失（分别在什么情况？）问题产生，用门（gate）控制每一state的信息进入量
- new memory都是根据之前state及input进行计算，但是**GRUs中有一个reset gate控制之前state的进入量**，而在LSTMs里没有类似gate；
- 产生新的state的方式不同，LSTMs有两个不同的gate，分别是forget gate (f gate)和input gate(i gate)，而**GRUs只有一种update gate(z gate)**；
- LSTMs对新产生的state可以通过**output gate(o gate)**进行调节，而GRUs对输出无任何调节。
- 最终GRU通过简化gate结构达到了相似的效果同时减少了很多参数与计算量

树模型

利用树模型 (lgb) 协助提升深度学习模型学习高频数据因子的能力

lgb和xgb异同和优劣？如何避免过拟合？

- 直方图优化，对连续特征进行分桶，在损失了一定精度的情况下大大提升了运行速度，并且在gbm的框架下，基学习器的“不精确”分箱反而增强了整体的泛化性能；
- (XGB也提出了类似的近似分位数算法，对特征进行分桶，即找到l个划分点，将位于相邻分位点之间的样本分在一个桶中。在遍历该特征的时候，只需要遍历各个分位点，从而计算最优划分。从算法伪代码中该流程还可以分为两种，全局的近似是在新生成一棵树之前就对各个特征计算分位点并划分样本，之后在每次分裂过程中都采用近似划分，而局部近似就是在具体的某一次分裂节点的过程中采用近似算法。重要的思想是：近似方法通过特征的分布，按照百分比（三分位数）确定一组候选分裂点，通过遍历所有的候选分裂点（遍历的范围缩小了）来找到最佳的分裂点。）
- 基于histogram决策树算法
- 直接支持对于类别特征进行训练（在对离散特征分裂时，每个取值都当作一个桶，分裂时的增益算的是“是否属于某个category”的gain。类似于one-hot编码）
- **树的生长方式由level-wise变成leaf-wise** 区别是xgboost对每一层所有节点做无差别分裂，可能有些节点的增益非常小，对结果影响不大，但是xgboost也进行了分裂，带来了务必要的开销。leaf-wise的做法是在当前所有叶子节点中选择分裂收益最大的节点进行分裂，如此递归进行，很明显leaf-wise这种做法容易过拟合，因为容易陷入比较高的深度中，因此需要对最大深度做限制，从而避免过拟合。

前沿

针对股票日品以及数字货币日频的因子数据的信号合成进行研究

介绍一下跟踪的深度学习在金融领域的新进展

对GNN有哪些了解，可能有哪些有效的应用方向

数字货币与传统金融标的的数据有哪些区别？针对这些区别是否有什么针对性的改进？