2 论文基本信息 EI检索

3 在训练集每次迭代过程中，产生一个新的子网络 Dropout的神经网络进行训练可以被看作是训练一系列具有大量权重共享的稀疏网络，如果有的话，每个稀疏网络得到非常少的训练。

当然在验证集时候稍有不同

4

模型组合几乎总能提高机器学习方法的性能。 然而，对于大型神经网络，平均许多单独训练网络输出的显而易见的想法是非常昂贵的。 当各个模型彼此不同时，结合几个模型是最有用的，为了使神经网络模型不同，它们应该有不同的体系结构或者受不同的数据训练。 训练许多不同的体系结构是很困难的，因为为每个体系结构寻找最优的超参数是一项艰巨的任务，并且训练每个大型网络需要大量计算。 而且，大型网络通常需要大量的训练数据，并且可能没有足够的数据在不同的数据子集上训练不同的网络。 即使有人能够训练许多不同的大型网络，在测试时使用它们都是不可行的，因为这对于快速响应非常重要。

Dropout 它可以防止过拟合，并且提供了一种有效结合指数级多种不同神经网络架构的方法。具有n个单位的神经网络可以看作是2n个可能的瘦神经网络的集合。 这些网络共享权重，因此参数总数仍然是O（n2）或更少。

Left: A unit at training time that is present with probability p and is connected to units in the next layer with weights w. Right: At test time, the unit is always present and the weights are multiplied by p. The output at test time is same as the expected output at training time.

这保证了期望输出和测试时候的真实输出是一样的。我们发现，训练一个dropout网络，并在测试时间使用这种近似平均方法，与其他正则化方法相比，可以在广泛的分类问题上显着降低泛化误差。

**2 动机 3 相关工作 4 描述dropout模型 5 训练dropout网络算法**

**6 实验结果 7分析在不同属性上使用dropout影响 并描述dropout与hyperparameters的相互作用 8 介绍Dropout RBM模型 9 探讨 marginalizing dropout**

2 MOTIVATION

动机1 进化中性别的作用 无性繁殖类似于创建父母基因略微变异的副本。可有幸繁殖是优越的 类似地，神经网络中的每个隐含单元在退出训练时都必须学会与随机选择的其他单元样本一起工作。这应该使每个隐藏的单元更加强大，并推动它自己创建有用的特性，而不依赖于其他隐藏单元来纠正其错误

3 Related work ignore

4. Model Description

每个神经元向后层传播时候添加一个独立的波努力随机变量

5 Learning Dropout Nets

5.1 随机梯度下降

可以使用随机梯度下降以类似于标准神经网络的方式来训练Dropout神经网络。 唯一不同的是，对于小批量中的每个train case，我们通过dropout单位来对稀疏网络进行抽样。 该train case的前向和后向传播仅在这个thinned的网络上完成。 在每个小批量的train case中，对每个参数的梯度进行平均。 任何不使用参数的train case都会为该参数贡献零的梯度。 许多方法已被用于改善随机梯度下降，如动量，退火学习率和L2权重衰减。 这些被发现对退出神经网络也很有用。

*我们发现一种特殊形式的正则化-dropout 将每个隐藏层单元的输入加权向量的范数限制在常数C，换句话说 如果w表示入射到任何隐藏单元上的权重向量，则神经网络是在约束|| w || 2≤c下优化 即 最大范数正则化 最大规范正则化过去曾用于协同过滤（Srebro和Shraibman，2005）。*

5.2 Unsupervised Pretraining

6. Experimental Results

**7. Salient Features**

我们看到dropout如何影响隐藏单元激活的稀疏性。*标准的神经网络中，每个参数会根据导数进行变化，来减少损失函数。所有的神经单元都在这样做，所以，一个神经单元的错误可能通过其它神经单元的修改来弥补。这就导致了复杂的联合适用，从而导致过拟合。Dropout可以利用不可靠性来防止过拟合。如图7.1*

Figure 7: Features learned on MNIST with one hidden layer autoencoders having 256 rectiﬁed linear units.

7.2对稀疏性的影响

7.3 Dropout Rate 影响

取值在0.4~0.8 图8说明 一般选取0.5效果最好

7.4 对数据集大小影响

对数据集大小影响不大

7.5

This suggests that the weight scaling method is a fairly good approximation of the true model average.

8 跳过

9.