在上一节介绍的遗传神经网络算法中，当误差逐渐减小到可接受的范围之后，或者训练次数达到人为设定的上限时停止训练。这样会产生一个问题，当训练数据集的误差持续降低，测试数据集的误差可能越来越大，会出现“过拟合”的现象。产生这种问题的原因在于，训练数据并不能完全代表测试数据集，即不能完全表达测试数据集的特征，所以如果训练数据量过小或者神经网络在训练过程中过度地学习训练数据中的细节和噪音，很可能导致训练结束后的BP神经网络对于训练数据的变现良好但是对于测试数据集的预测判断能力不尽如人意，这时该网络的泛化能力较差，所以网络用于实际重复记录检测时效果也较差。如下图所示：

  

(a)欠拟合 (b)理想拟合 (c)过拟合

图4.5 不同拟合情况示意图

图中蓝色的函数曲线代表的模型是对黄色方块从X到Y的映射模拟，图(a)展示的是一种“欠拟合”，即训练好的模型对于训练数据的表现能力很差，所以其对测试数据的表现一定也很差；图(b)展示了一种合适的拟合，对训练集和测试集的表现均良好，是一种最理想的拟合模型；图(c)则展示了“过拟合”的情况，虽然从图中可以看出该模型对训练数据集表现接近完美，但是很明显该模型用于测试数据集时表现会很差，这就是过度拟合导致的泛化能力的下降。

为了防止这种过拟合的出现，或者缓解其带来的消极影响，本文将“Dropout”方法应用到此BP神经网络中，以提高其泛化能力。

“Dropout”[[[1]](#endnote-1)]由多伦多大学的Hinton教授于2014年提出，它是一种神经网络训练过程中的技巧。Dropout是指在训练过程中，按照一定的概率暂时关闭神经网络中的神经元，使其不发挥作用，从而实现避免神经网络过拟合的目的。

如下图所示：

 

(a)标准神经网络 (b)应用dropout后的神经网络

图4.5 dropout网络结构示意图

由(a)图和(b)图的对比可以看出，dropout相当于减弱了原网络的复杂度，暂时得到了一个更简单的结构更稀疏的神经网络。

在训练过程中，采用dropout相当于对每一个结点以一定的概率使其关闭，则结点发挥作用的概率为，所以对于神经网络中的每一个结点，其发挥作用与否服从伯努利分布。应用dropout前后，神经网络的计算方式区别如下图所示：

 

(a)标准神经网络 (b)应用dropout后的神经网路

图4.6 应用dropout计算方式区别示意图

在标准的神经网络结构下，训练时的计算方法如下：

 公式 4-7

即为输入和权值的乘积和再加上偏置，然后应用激活函数得到输出。当应用了dropout之后，上面的神经网络的计算方法如下：

 公式 4-8

即首先要对输入乘以一个按照概率服从伯努利分布的随机变量，一般的取值为0.5。

Dropout能够避免过拟合的原因主要有以下两点：

（1）减弱神经元之间的共适应关系。Dropout操作导致神经元之间的结构关系更加动态化，这样不同的神经元之间不一定每次都能同时出现在网络中，这种情况下对权值进行更新则不依赖于固定的网络结点的相互作用，这就强迫神经网络去学习更加鲁邦的特征。也就是说，这种情况下神经网络不会对特定的线索过于敏感，从而能一定程度上减弱神经网络对于某些特殊情况（错误值、缺失值等）的过于拟合。

（2）Dropout起到了取平均值的作用。不同的网络可能产生不同的过拟合，采用dropout则会产生不同的神经网络，这些网络中存在的一些互为“反向”的过拟合在综合起来取平均的作用下可互相抵消，这就在整体上减少了过拟合。

将dropout应用于BP神经网络的训练过程，算法的流程图如下：

1. [] Hinton 提出了一个神器，

   Dropout: A Simple Way to Prevent Neural Networks from Overfitting [↑](#endnote-ref-1)