4.2 基于BP神经网络的赤潮预测模型

Forecasting model based on BP neural network

4.21输入输出因子的选取

The selection of output and input factors

由于研究对象的特殊性，我们只能通过各类网站及报告获取到零碎的数据，经过数据规整化后，我们获得了129组数据，由于水体叶绿素a浓度是表征水体中藻类现存量的最直接指标，故水体中总叶绿素a浓度是模型的输出因子，通过对叶绿素a浓度的预测可以间接实现对藻类引发的水华进行预测．.在确定模型输出因子后，合理选择网络的输入因子对正确应用BP模型和保证模型预报精度非常重要。因为输入因子中，可能存在与输出因子关联弱的噪声因子，或者重叠反映系统信息的冗余因子．无论是噪音因子还是冗余因子，都会加大分析问题的难度，增加模型的复杂性，最终影响模型预测能力．筛选BP神经网络输入因子的基本原则是选择与输出因子相关而又彼此无关的环境因子网络结构。

In view of particularity of studying object, we get 129 groups data processed which are from all kinds of web sites and papers. The reason why the potentially-toxic algal blooms reproduction can be predicted indirectly via forecasting the content of [chlorophyll](http://cn.bing.com/dict/clientsearch?mkt=zh-CN&setLang=zh&form=BDVEHC&ClientVer=BDDTV3.5.0.4311&q=%E5%8F%B6%E7%BB%BF%E7%B4%A0a" \t "_blank) is that the total content of [chlorophyll](http://cn.bing.com/dict/clientsearch?mkt=zh-CN&setLang=zh&form=BDVEHC&ClientVer=BDDTV3.5.0.4311&q=%E5%8F%B6%E7%BB%BF%E7%B4%A0a) [a](http://cn.bing.com/dict/clientsearch?mkt=zh-CN&setLang=zh&form=BDVEHC&ClientVer=BDDTV3.5.0.4311&q=%E5%8F%B6%E7%BB%BF%E7%B4%A0a) is the most direct index to represent the quantity of algal blooms. After determining the output factor, reasonable selection of network‘s input factors is of great importance to apply BP neural network models accurately and guarantee the precision of the model, for input factor may exist noise factor related to outputs and redundant factor reflecting system information. Either of the two conditions will add [complexity](http://cn.bing.com/dict/clientsearch?mkt=zh-CN&setLang=zh&form=BDVEHC&ClientVer=BDDTV3.5.0.4311&q=%E5%A2%9E%E5%8A%A0%E6%A8%A1%E5%9E%8B%E7%9A%84%E5%A4%8D%E6%9D%82%E6%80%A7) to model and difficulty to analyze problem so as to affect the [predictive](http://cn.bing.com/dict/clientsearch?mkt=zh-CN&setLang=zh&form=BDVEHC&ClientVer=BDDTV3.5.0.4311&q=%E9%A2%84%E6%B5%8B%E8%83%BD%E5%8A%9B) [ability](http://cn.bing.com/dict/clientsearch?mkt=zh-CN&setLang=zh&form=BDVEHC&ClientVer=BDDTV3.5.0.4311&q=%E9%A2%84%E6%B5%8B%E8%83%BD%E5%8A%9B) of model. The principle of sorting out input factors of neural network is to select environment factors of network structure which have nothing to with outputs.

最终，我们以N,P输入总量作为输入因子，去除了数据量过少的因子以及其他可能的噪声因子及冗余因子。

At last, we regard N,P input totals as input factors and get rid of factor lacking data and noise factor and redundant factor.

4.22网络结构

Determine Network structure

早有理论证明：3层BP网络，当各层神经元均采用S型函数时，可满足任意复杂的非线性函数拟合逼近问题【7】．

[Theoretical](http://cn.bing.com/dict/clientsearch?mkt=zh-CN&setLang=zh&form=BDVEHC&ClientVer=BDDTV3.5.0.4311&q=%E7%90%86%E8%AE%BA%E8%AF%81%E6%98%8E" \t "_blank) [proof](http://cn.bing.com/dict/clientsearch?mkt=zh-CN&setLang=zh&form=BDVEHC&ClientVer=BDDTV3.5.0.4311&q=%E7%90%86%E8%AE%BA%E8%AF%81%E6%98%8E" \t "_blank) long before showed that, three BP layers can satisfy any complicated **nonlinear function to fit** [approximation](http://cn.bing.com/dict/clientsearch?mkt=zh-CN&setLang=zh&form=BDVEHC&ClientVer=BDDTV3.5.0.4311&q=%E9%80%BC%E8%BF%91%E9%97%AE%E9%A2%98) [problem](http://cn.bing.com/dict/clientsearch?mkt=zh-CN&setLang=zh&form=BDVEHC&ClientVer=BDDTV3.5.0.4311&q=%E9%80%BC%E8%BF%91%E9%97%AE%E9%A2%98" \t "_blank)when every layer neurons adopt **sigmoid function. The conclusion can provide reference for ensuring** algal blooms as three layers network structure, one input layer, one hidden layer and one output layer. Although selecting the number of hidden layer’s neurons and activation function between layer and interlayer is regular, the results are even altogether different.

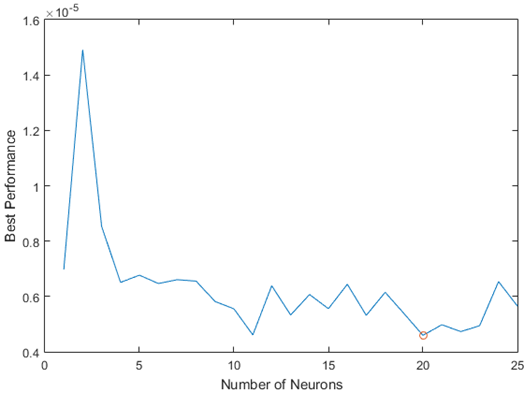
这个存在性结论对神经网络结构的设计具有重要的指导作用．参考该结论，确定水华预测模型为3层网络结构，即1个输入层、1个隐含层和1个输出层．然而，选取隐含层神经元数、层与层间的激活函数，虽然有规则可依，但所得结果差异较大，甚至完全不同【8】．

4.23隐含层神经元数

Select neurons of hidden layer

通过对不同隐含层神经元数的模型进行训练，使用Mean squared normalized error performance function来分析隐含层神经元数对训练效果的影响，从分析结果发现：当隐含层神经元数大于3个时，训练效果开始随隐含层神经元数的增长在小范围波动，总体上效果很好，当隐含层神经元数为20时，得到最好结果,加上我们的数据集不是很大，因此我们以20作为隐含层神经元数.

Through training neurons of different hidden layers, using mean squared normalized error performance function analyzes the numbers of neurons in hidden layer on the impacts of training effects, then we draw a conclusion from results that training effects [fluctuate](http://cn.bing.com/dict/clientsearch?mkt=zh-CN&setLang=zh&form=BDVEHC&ClientVer=BDDTV3.5.0.4311&q=fluctuate" \t "_blank) in a small scale with the increasement of neurons of hidden layer in number when they are more than three. Moreover, we can get the best result when the numbers of neurons are twenty, so we set up twenty neurons as hidden layer considering not many data.



4.24激活函数

**Activation Function**

We set up a sigmoid transfer function in the hidden layer and a linear transfer function in the output layer

依照上述的网络结构，得到网络结构拓扑以及模型表达式According to the above network structure ,we have network topology and

y=purelin(V·tansig(W·x+b1)+b2)．

$$y=purelin(V\**cdot** tansig(W\**cdot** x+b\_**1**)+b\_**2**)$$

式中：x为输入变量， Y为输出变量，即当前N，P浓度下的叶绿素a浓度；w和b。为输入层与隐含层问连接权重和阈值；y和b。隐含层与输出层间的连接权重和阈值。

In the equation:x represents N,P concentrations and y represents the content of [chlorophyll](http://cn.bing.com/dict/clientsearch?mkt=zh-CN&setLang=zh&form=BDVEHC&ClientVer=BDDTV3.5.0.4311&q=%E5%8F%B6%E7%BB%BF%E7%B4%A0a" \t "_blank) a ,w and b represent link weights between input layer and hidden layer. Y and b represent link weights between hidden layers and output layer.