第三届“泰迪杯”

全国大学生数据挖掘竞赛

优

秀

作

品

作品名称：城市供水处理混凝投药过程的建模与控制

荣获奖项：二等奖

作品单位：湖北工程学院

作品成员：贾园园 万爽 裴幸智

指导教师：张学新

基于 BP 神经网络的最佳投药量预测

摘　要： 混凝投药通过投加混凝剂除去原水中的杂质及其他有害物质，是城市供水过程中的重要环节之一，这一过程效果的好坏将直接影响后续处理工艺及出水水质的好坏。该过程具有影响因素多、大滞后性和非线性等特征，实际控制难度较大。本文基于广州南沙水厂提供的 9397 个投药控制数据，尝试构建一种基于 BP 神经网络混凝投药控制模型，来预测混凝剂的最佳投药量。

对于原数据集有缺失值情况，本文做基本预处理，用三次样条插值法对出水浊度进行插值估算，并剔除 5　外的极端异常值，按照出水浊度小于 1.10NTU 的标准，筛选出投药合格的 6143 个数据，以此作为样本数据。

针对第（1）问，本文运用平流沉淀理论，求得原水混凝沉淀到出水结束的滞后时间，约为 80 分钟，在实际范围 70min--120min 内。

针对第（2）问，本文以原水浊度、原水流速、原水 PH 值三个因素作为 BP 神经网络模型的输入神经元参数，对混凝剂投加量的训练样本和测试样本进行分析，得到预测的最佳投药量；

针对第（3）问，在第二问之上，增加出水浊度做为输入参数再次建立 BP 神经网络模型，并与第（2）问的模型进行比较。为了比较模型性能，我们又建立多元线性回归模型，找出四个变量与投药量的回归方程，通过在训练样本与测试样本上的预测效果，对 BP 神经网络模型和多元回归模型进行比较，分析绝对误差等指标，发现 BP 神经网络具有更强的非线性逼近能力，能够对投药量进行很好的仿真和预测效果。

针对第（4）问，本文查找文献[8]，引入温度数据，验证文献[9]的理论模型，通过对数变换化为线性模型，并对模型的整体显著性和温度系数的显著性作检验，但是最后结果表明系数的显著性并不强，即温度对投药量的影响并不大，并从有关化学理论角度对此结果进行解释。

关键词： 混凝投药；平流沉淀理论；BP 神经网络；多元线性回归；最佳投药量

The optimal dosage prediction based on BP neural network

Abstract: Coagulation Dosing remove impurities by adding coagulant and other harmful substances in the raw water, is an important part of the city water supply process, this process is good or bad will directly affect the subsequent treatment process and water quality is good or bad. This process has many factors, large hysteresis and non-linear characteristics, the actual control more difficult. 9397 Dosing Control Based on data provided by Guangzhou Nansha water, try to build a BP neural network based on the optimal dose administered Coagulation Dosage Control model to predict coagulant.

For the original data set has missing values, this paper do basic pretreatment of water turbidity estimate interpolating cubic spline interpolation, and excluding extreme outliers 5 outside, according to the turbidity of less than 1.10NTU criteria selected Dosing qualified 6143 data, as sample data.

For the first (1) asked this paper, the theory of advection precipitation, coagulation and sedimentation to obtain raw water outlet end of the lag time of about 80 minutes, the actual scope of 70min - within 120min.

For the first (2) Q, the paper raw water turbidity raw water flow rate, the raw water PH value of the three factors as the input neuron parameters BP neural network model, the training and testing samples coagulant dosage was analyzed, Best predicted dose administered;

For the first (3) asked on the second question, increased water turbidity as input parameters to establish BP neural network model again, and with paragraph (2) Q models were compared. To compare the performance of the model, we have set up multiple linear regression model to identify the four variables and dosage of the regression equation, by predicting the effect on training samples and testing samples of BP neural network model and multiple regression model comparison and analysis absolute error and other indicators, found that BP neural network has stronger nonlinear approximation ability to perform well on the dosage of the simulation and prediction.

For the first (4) Q, the paper find [9], the introduction temperature data validation [10] A theoretical model by logarithmic transformation Huawei linear model, and the model of the overall significance and temperature coefficient of significance for examination , but the final results show a significant coefficient is not strong, the effect of temperature on the dosage

amount is not large, and from about the stoichiometric point of view to explain this result.

Key words: Coagulation administration; Advection precipitation theory; BP neural network; Multiple linear regression; The optimum dosage

目 录

1. 挖掘目标

针对现有水处理混凝投药控制方法中的不足，如烧杯实验耗时多；流动电流法精度低，适用范围狭窄；数学模型法不能适应控制情况而变化等问题。

本次建模的目标是利用广州南沙水厂投药控制系统实时采集的数据信息建立基于BP神经网络的混凝投药控制模型，此模型能够很好的对投药控制的各种不定因素进行非线性逼近，而且可以随着条件的需要变换修正各项参数，解决混凝投加与对应水絮凝沉淀后的浊度存在一段较长的时间差所造成的控制滞后问题，从而很好的预测出最佳混凝剂投药量，避免不必要的浪费，在保证水质的前提下节约水厂制水成本。

2.分析方法与过程

☆2.1 总体流程

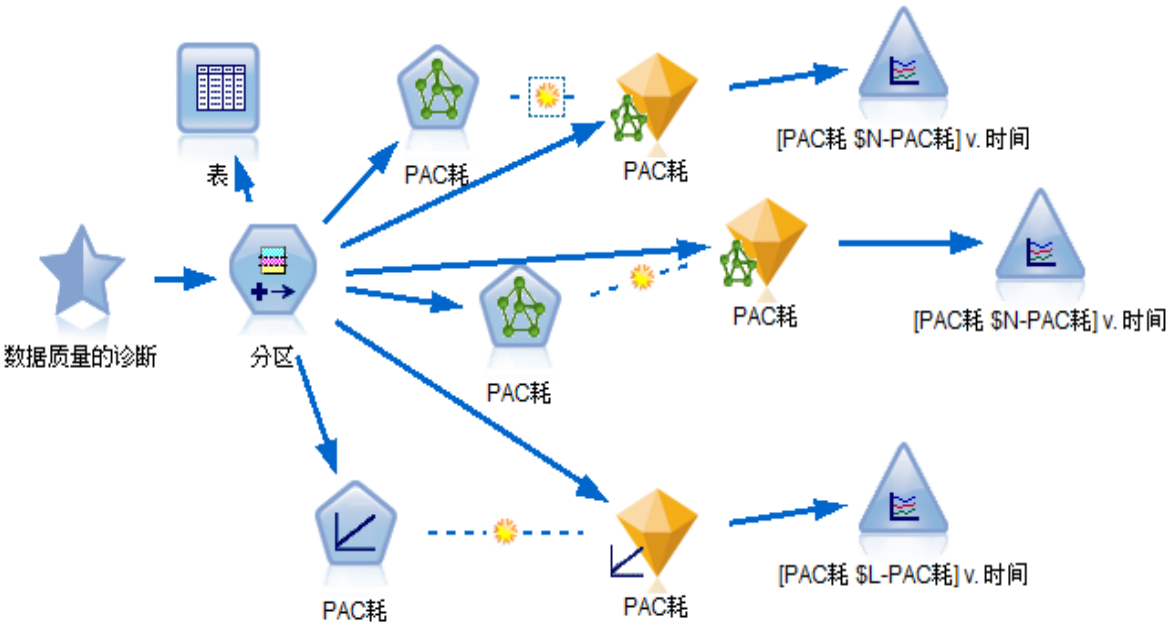


图 1.0 总数据流图

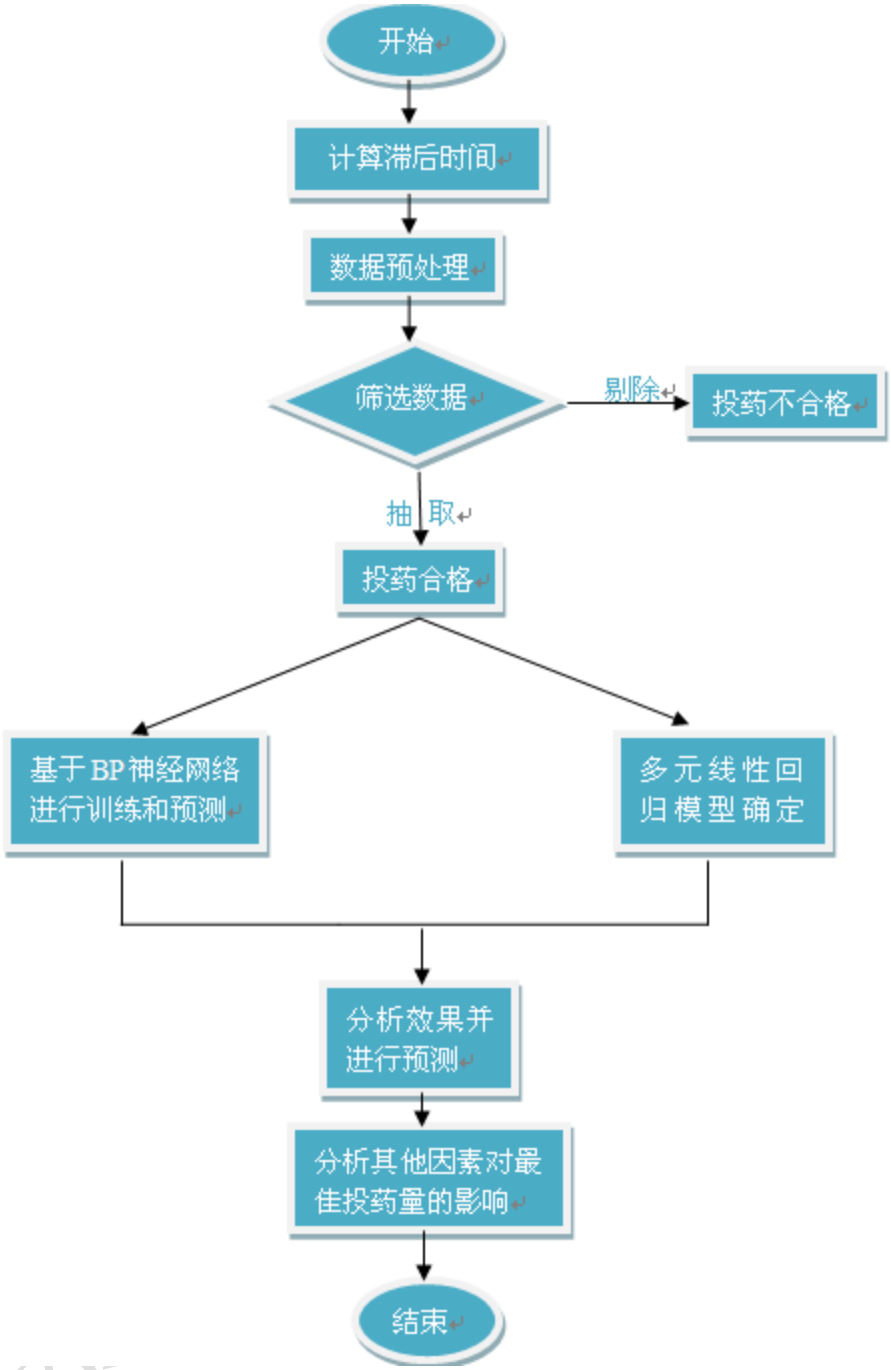


图 2 结构流程图

混凝投药过程包括以下几个步骤：

步骤一：结合已给样本数据集求出原水添加混凝剂反应到沉淀结束出水的滞后时间。

步骤二：对数据进行预处理。

步骤三：由出水是否合格结合滞后时间反推投药是否合格，筛选出所有投药合格的数据。

步骤四：基于 BP 神经网络构建模型，对数据进行训练与预测。

步骤五：构建多元线性回归模型确定表达式。

步骤六：分析两种模型的仿真和预测效果并进行比较，给出最优模型。

☆ 2.2 具体步骤

▪2.2.1 计算滞后时间【问题（一）求解】

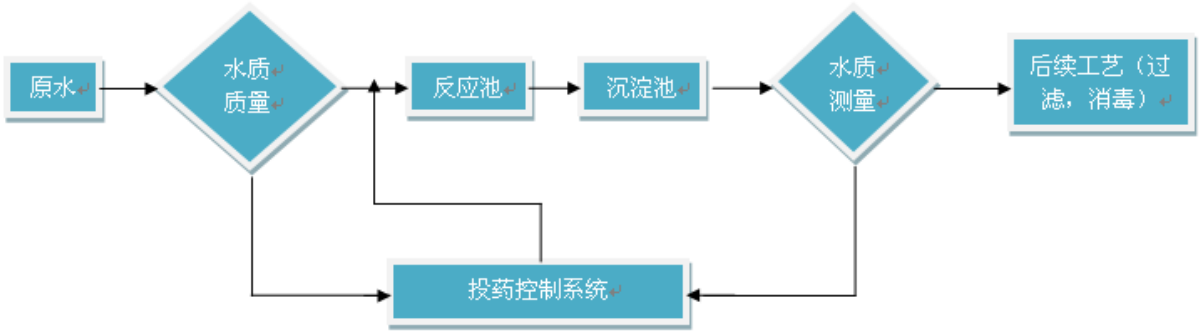


图 3 投药控制流程图

如图 3 投药控制流程图，由于混凝沉淀池是一个大容积对象，对于混凝剂投加与对应水絮凝沉淀后的浊度存在一段较长的时间差。滞后性是水处理过程中的一大难题，因此，求出有效地滞后时间显得尤为重要。

根据有关资料，了解到平流沉淀池的理论停留时间为[1]：

$$T = \frac{V}{Q} \tag{1}$$

其中V表示沉淀池的有效容积，Q表示原水平均流量。

为此，我们调查了广州南沙水厂混合反应沉淀池的相关数据[2]，其工艺采用竖流式折板絮凝池。整个絮凝池设计分为三段，每段絮凝池容积依次递增，絮凝区总反应时间为 15.0min。平流沉淀池 4 座，每座分两格，单座尺寸为$78.5m \times 28.8m \times 5.1m = 11530.08m^3$，计算得到水平流速为$8651.707m^3/h$，有效水深为$4.0m$，因此理论时间为：

$$T = \frac{78.5 \times 28.8 \times 4.0}{8651.707} \approx 1.08h \tag{2}$$

综上，总滞后时间等于反应时间加沉淀时间，即为$1.33h$，约为 80 分钟。

▪2.2.2 数据预处理

原始数据中有很多缺失值和异常值，如果不对数据进行预处理而直接使用，会给今后的工作造成很大的困扰。为了确保辨识模型参数的准确性、可靠性，需要对原始数据进行预处理，尽可能减小系统误差和随机误差对模型辨识的干扰，图 4 和图 5 所示为混凝投药过程变量的原始数据与处理后数据（对比图见附录 1 图表附件），横坐标表示数据个数，纵坐标表示实际测量值，图形有些地方跳动非常大，这一些跳动的地方就代表异常数据。如果直接把这些数据用于过程模型辨识，辨识出的效果肯定不是特别理想，甚至会与理论模型相差很大，对后续的工作造成很大的困扰，因此在辨识阶段剔除这些异常数据显得非常必要。

样本数据预处理包括缺失值处理、离群值处理、极端值处理[3]。

- 缺失值处理

在原始数据中，发现 8 月 20 日之前的 PAC 耗存在缺失的现象，大约占总数据的 3%，为确保建模数据的有效性，这里将异常值剔除，不计入后面的计算当中。

- 极端值处理

本文将距离均值 5σ 以外的值称为极端值，极端值对数据的影响较大，严重超出指标范围，因此对于极端值，同样要将其剔除。

- 离群值处理

这里将距离均值 3σ 以外的值称为离群值，离群值不宜剔除，因此将超出 3σ 的值取作 $\mu \pm 3\sigma$ 的端点值。

本文使用 IBM SPSS MODELER 14.1 对所有数据进行数据质量检测，包括缺失值、极端异常值处理，得到处理前与处理后的一些数据表示如下：

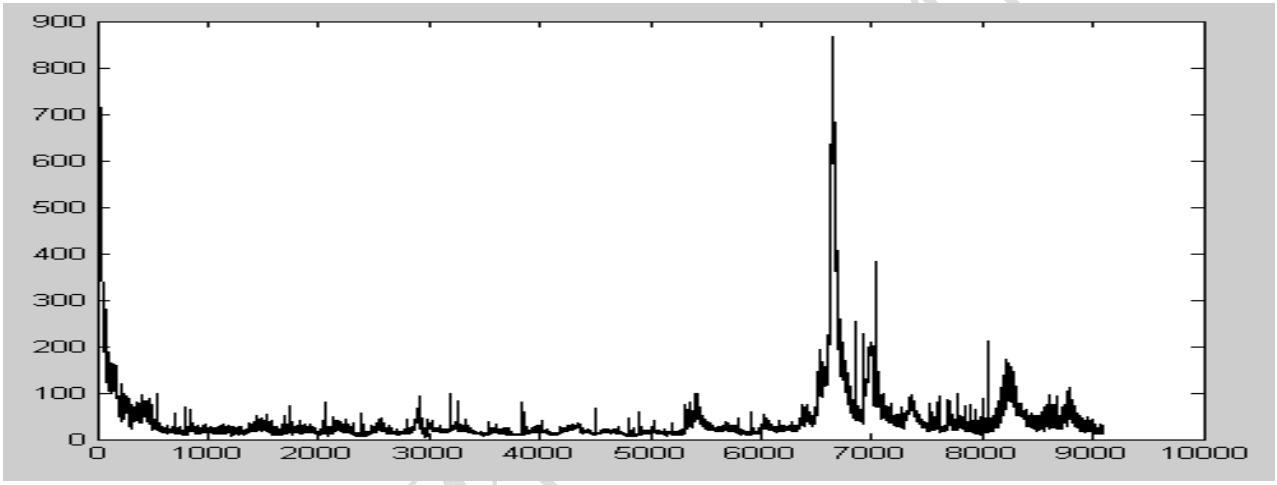


图 4 原水浊度原始数据

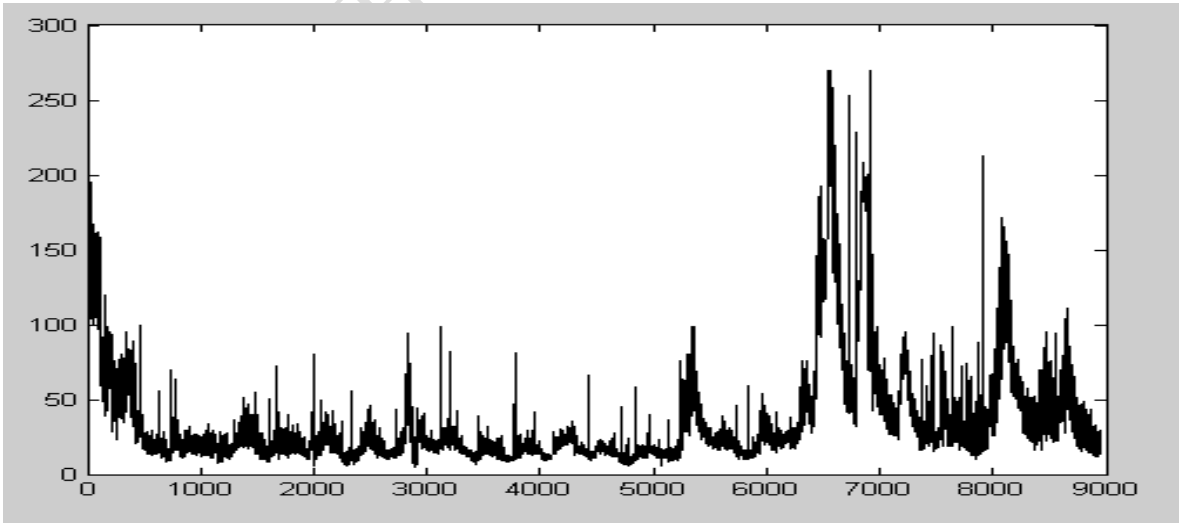


图 5 处理后原水浊度数据

由前后对比，可以明显看出，处理后的数据质量明显优于处理前，样本极端值和异常值均已被处理，处理后的各项指标的标准差和偏度都小于处理前，数据的稳定性得到大幅提高，这为之后的建模过程打下坚实的基础。

▪2.2.3 筛选数据

按题目要求将 3 号池和 4 号池的平均浊度作为出水浊度。

由于所算出的滞后时间为 80 分钟，需要将出水浊度数据的频率由 1 小时细分为 20 分钟，即需要在每两次测量之间插入两个值。数据插值的常见方法主要有以下几种：三次样条插值、分段三次 Hermite 插值、分段线性插值、Hermite 插值[4]。

本文用三次样条插值法对出水浊度进行插值估算，并运用 MATLAB 来实现此过程，程序见（附件 2 插值程序）。

若出水浊度小于等于 1.10NTU，在 Excel 中取 0，说明出水合格，若出水浊度大于 1.10NTU，则在 Excel 中取 1，说明出水不合格。

由出水是否合格反推出投药是否合格，如图表示投药合格与不合格分别所占比例如图 6：

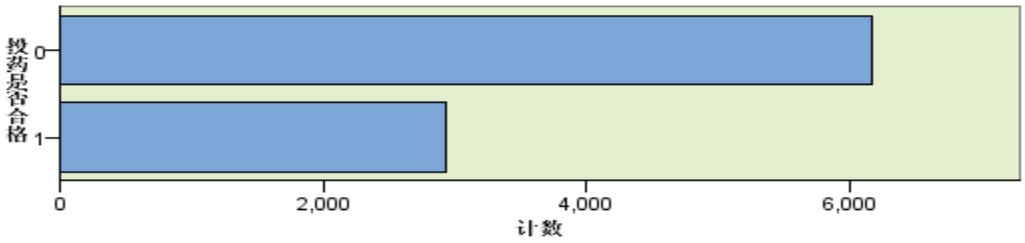


图 6 投药合格比例

投药合格有 6143 个数据，占总体比例为 68.59%；投药不合格有 2813 个数据，占总体比例为 31.41%。随后本文将采用所有投药合格的数据用于 BP 神经网络模型及多元回归模型的仿真及预测。

▪2.2.4 构建 BP 神经网络模型【问题（2）（3）对比求解】

误差反向传播神经网络，简称 BP 神经网络，是一种单向传播的多层前向网络，它是一种具有自适应能力的高度非线性的动力学系统，能对未知非线性函数具有出色的逼近与学习功能，通过对人脑的形象思维、联想记忆等功能进行模拟、简化和理论抽象，可以用来描述认知、决策和控制等智能行为。神经网络在结构上具有并行处理、分布式存储和容错性强等特征，在能力上具有自学习、自组织和自适应等特征，使得神经网络具备处理不确定的、非结构化的信息的能力[5]。

2.2.4.1 BP 神经网络结构

图 7 给出了一个具有 r 个输入和 BP 神经网络结构模型，每个输入都通过一个适当的权值和下一层相连，网络的输出可表示为： $a=f(wp+b)$， f 表示输入、输出关系的传递函数[6]。

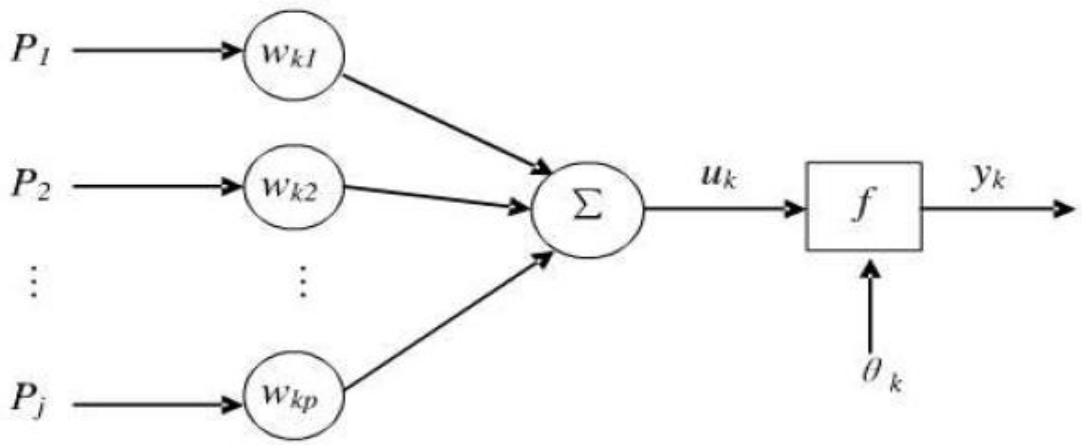


图 7 基本神经元模型

2.2.4.2 BP 神经网络建模流程

BP 神经网络的混凝投药预测建模流程主要包括：

（1）确定网络层数

现已经证明三层 BP 网络可以逼近任意非线性特性，因此本文所建立的混凝投药预测模型神经网络层数为 3。

（2）确定输入层的层数

根据题目的第 2、3 问，先取原水 PH、原水浊度、原水流量三个变量作为输入层，建立 BP 神经模型，求出最佳投药量预测值；然后再增加沉淀池出水浊度作为输入参数共四个变量构建 BP 神经网络数学模型求最佳投药量。如图 8,9 所示：

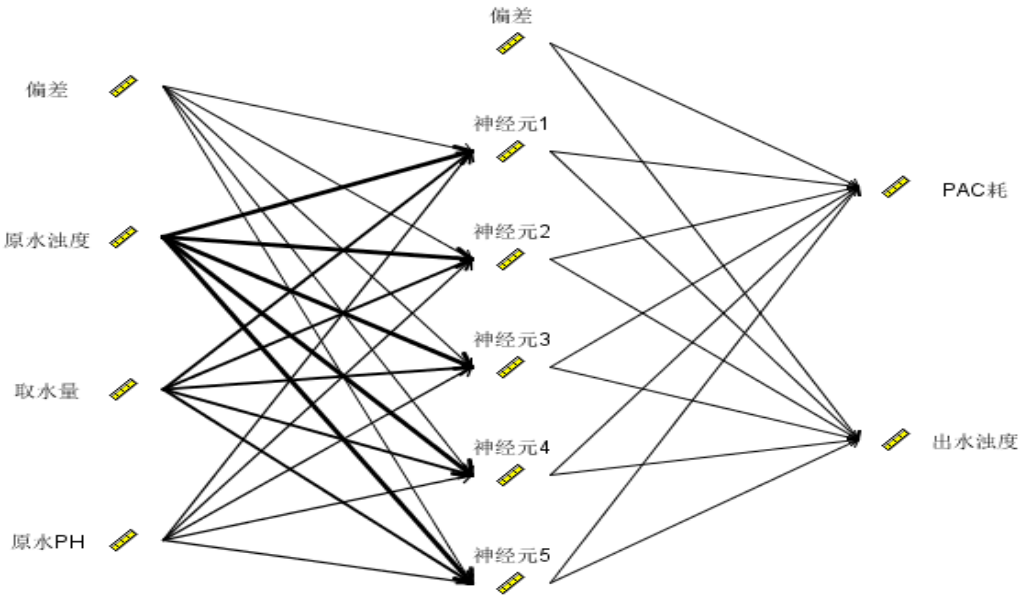


图 8 三变量神经网络模型结构

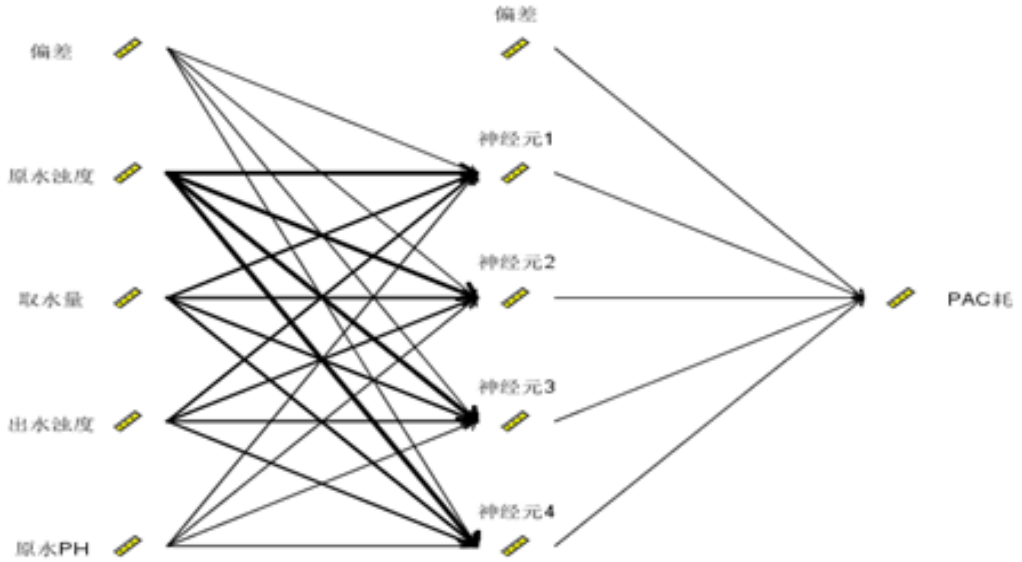


图 9 四变量神经网络模式结构

（3）数据预处理

各个输入数据具有不同的物理意义和不同的量纲，BP 神经网络的神经元都是基于 Sigmoid 转移函数，变换后可以消除因输入变量绝对值过大而使得神经元的输出饱和，经过权值调整继而使得误差曲面变得平坦。Sigmoid 转移函数的值域规定在 0 到 1 或-1 到 1 之间，如果不进行预处理，数值大的分量和数值小的分量绝对误差相差非常大，但是网络训练只是针对系统输出的总体误差相应的调整权值。基于上述原因，大多数神经网络在运行之前，都需要提前对实际数据进行归一化处理，而且不能在神经网络内部进行，必须放在预处理阶段进行，这样做的好处是能显著减少神经网络的代码，从而其可移植性、通用性优于一般的算法。具体做法可以采用尺度变换解决这个问题，在整个范围内确定最小值 $x_{\min}$ 。和最大值 $x_{\max}$ 。，进行统一的变换处理。[0,1]区间数据变换的变换公式为：

$$\overline{x} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \qquad (3)$$

其中 x_i 表示输入或输出数据， $x_{\min}$ 表示数据变化的最小值， $x_{\max}$ 表示数据变化的最大值。

（4）确定隐节点数

通常，对于波动次数多、幅度变化大的复杂非线性函数，若要增强神经网络的映射能力，就要求具有较多的隐节点。一个常用的确定最佳隐节点的方法是试凑法。试凑法确定隐节点数的一般步骤为：首先设置少量隐节点用于训练网络；然后逐渐增加隐节点数，并保证训练的样本集不发生改变；最后通过不断调整和训练确定最佳隐节点数。

（5）训练神经网络

网络性能的好坏主要看其是否具有很好的泛化能力，通常将采集到的数据随机分成两部分，一

部分用作训练集，另一部分作为测试集，测试神经网络的泛化能力。这里将 70% 的数据作为训练样本，30%的数据作为预测样本。对三个输入变量与四个输入变量的模型分析结果做了对比,对比结果见表 1 ：

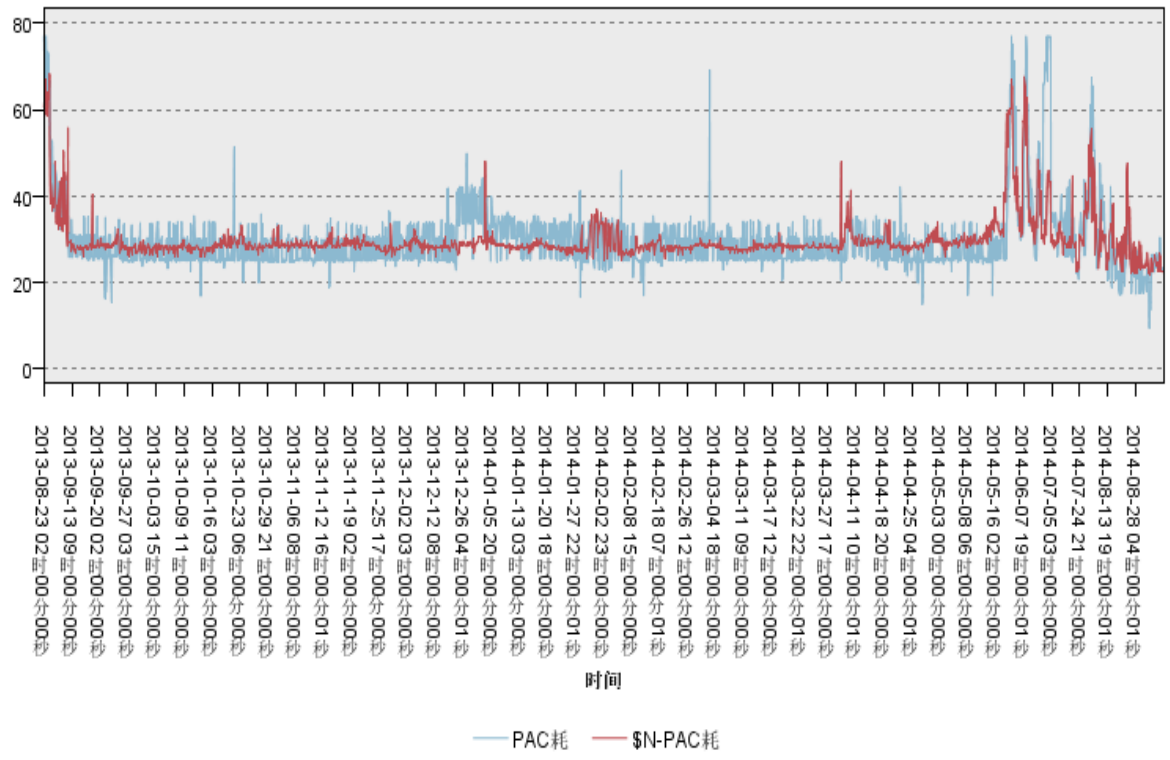
模型概要

	三变量	四变量
目标	PAC 耗	PAC 耗
模型	多层感知器	多层感知器
所使用的停止规则	无法进一步降低误差	无法进一步降低误差
隐藏层 1 神经元	4	4
准确度	44. 60%	51%

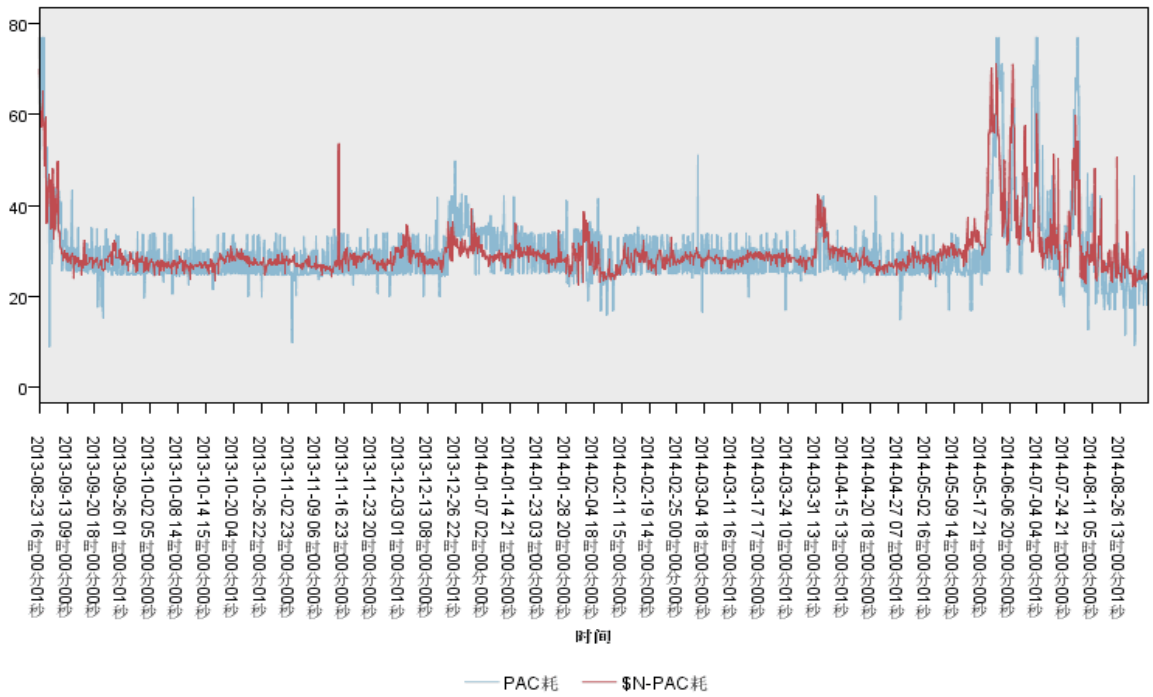
表 1.模型概要

表 1.可以看出四变量的准确度大于三变量的准确度，但是本次建模的准确度只有 44.6%和 51.0%，可以说这个估计效果并不是很好，可能是由于输入变量较少的原因。

图 10 反应了两次建模 PAC 耗的预测值与原始数据的对比曲线，可以看出，预测值十分稳定，起伏相对较小，对原始值的起伏趋势也很好的进行了拟合，模型训练及测试效果比较好。同样从两曲线也可看出四变量的训练效果及测试效果优于三变量。



三变量 PAC 耗预测与历史对比



四变量 PAC 耗预测与历史对比

图 10 PAC 耗与原始数据对比

表 2.相关误差分析误差

比较$N-PAC 耗与 PAC 耗

	三变量		四变量	
分区	1-训练	2-测试	1-训练	2-测试
最小误差	-36.935	-33.341	-36.082	-30.896
最大误差	42.709	44.21	42.049	39.581
平均误差	-0.103	-0.237	-0.001	-0.071
绝对平均误差	4.256	4.202	4	3.964
标准差	6.028	5.931	5.671	5.636
线性相关	0.668	0.65	0.714	0.692
发生率	4274	1871	4274	1871

表2.表示两次建模训练和测试的相关误差与相关性，同训练平均绝对误差相比，预测平均绝对误差较小，因此模型的预测效果要优于训练效果，但与原始数据的线性相关程度较弱一些。这里也可看出四变量的相对误差小于三变量模型的相对误差。

▪2.2.5 构建多元线性回归模型

为了和BP神经网络模型进行比较，我们运用相同的实验数据建立了多元线性回归的数学模型。模型自变量为原水PH、原水浊度，原水流量，出水浊度，数学模型的参数估计采用最小二乘法计算。利用SPSS MODELER软件进行参数求解和模型检验，以下对模型的仿真和预测过程进行简单的说明。

多元线性回归分析是以多个解释变量的给定值为条件的回归分析，是研究一个因变量和多个自

变量间的线性关系方法。这里我们建立起以原水浊度、流量、PH、出水浊度为自变量，投药量为因变量的多元线性回归模型：

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4 \quad (4)$$

其中，y ———混凝剂投加量；

β_i——回归系数；

x_1——原水PH；

x_2——原水浊度；

x_3——原水流量；

x_4——出水浊度。

多元线性回归模型的参数估计出来后，即求出样本回归函数后，还需要进一步对该样本回归函数进行统计检验，以判定估计的可靠程度，包括拟合优度检验（可决系数）、方程总体线性的显著性检验（F检验）、变量的显著性检验（t检验）。

（1）拟合优度检验（可决系数），用统计量来衡量样本回归观测值得拟合程度，记 $TTS=\sum(Y_i-\overline{Y})^2$ 为总离差平方和，$ESS=\sum(\hat{Y}-\overline{Y})^2$ 为回归平方和，$RSS=\sum(Y_i-\hat{Y}_i)^2$ 为残差平方和，则：

$$R^2=\frac{ESS}{TSS}=1-\frac{RSS}{TSS} \quad (5)$$

其中：R^2 为可决系数，该统计量越接近1，模型的拟合优度越高。

（2）方程总体线性的显著性检验（F检验），在原假设 H_0 成立的条件下，统计量 ：

$$F=\frac{ESS/k}{RSS/(n-k-1)} \quad (6)$$

服从自由度 $(k,n-k-1)$ 的F分布。因此，给定显著水平 a，查表得到临界值 $F_\alpha(k,n-k-1)$，根据样本求出F统计量的数值，通过 $F>F_\alpha(k,n-k-1)$ 来拒绝（或接受）原假设 H_0，以判断原方程总体上的线性关系是否显著成立[7]。

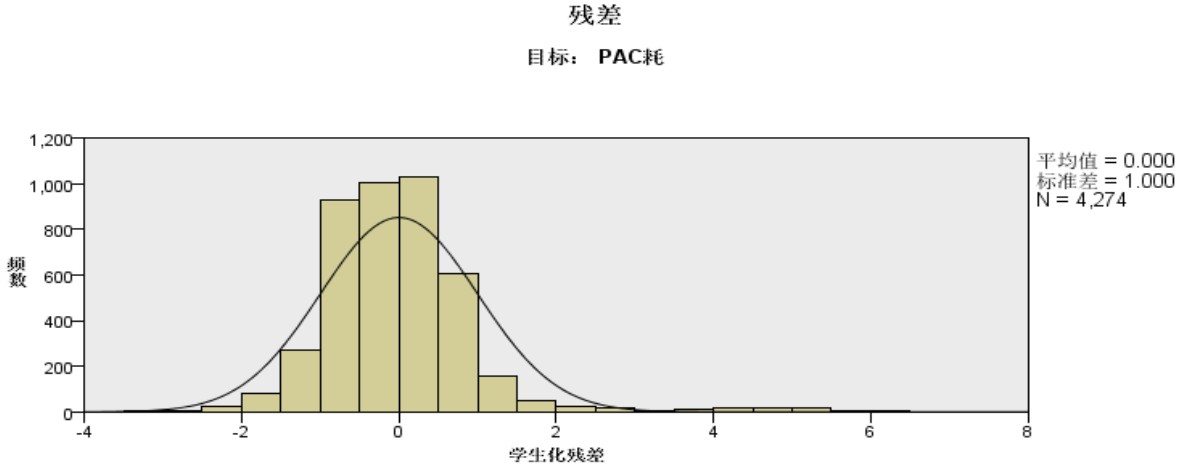


图11 ：残差分布图

IBM spss modeler 软件中可直接通过  图标来求解多元线性回归方程的各项系数。

将图11 残差分布图与正太分布曲线进行比较，发现残差频率比较靠近正态分布曲线，所以残差分布近似于正态分布。

回归模型的系数b确定之后，用多元线性回归方法建立的投加量预测数学模型可确定为：

$$y = 55.549 - 1.105x_1 + 0.204x_2 - 0.001x_3 + 10.595x_4$$

模型确立后，从方差分析上知道模型是有效可行的，为了更直观的表现其有效性，我们用原始建模数据对模型进行训练效果分析，图12是实际投药量与仿真投药量的对比曲线。从图中我们可以看出，仿真投药量与实际投药量的总体曲线走向是一致的，只有在实际投药量变化剧烈的时候，仿真投药曲线平滑过渡过去，造成模型计算的投药量与实际投药量的偏差比较大，但是由图11 残差分布可知，这些点占得比例很小，因此整个模型的训练效果还是比较好的。

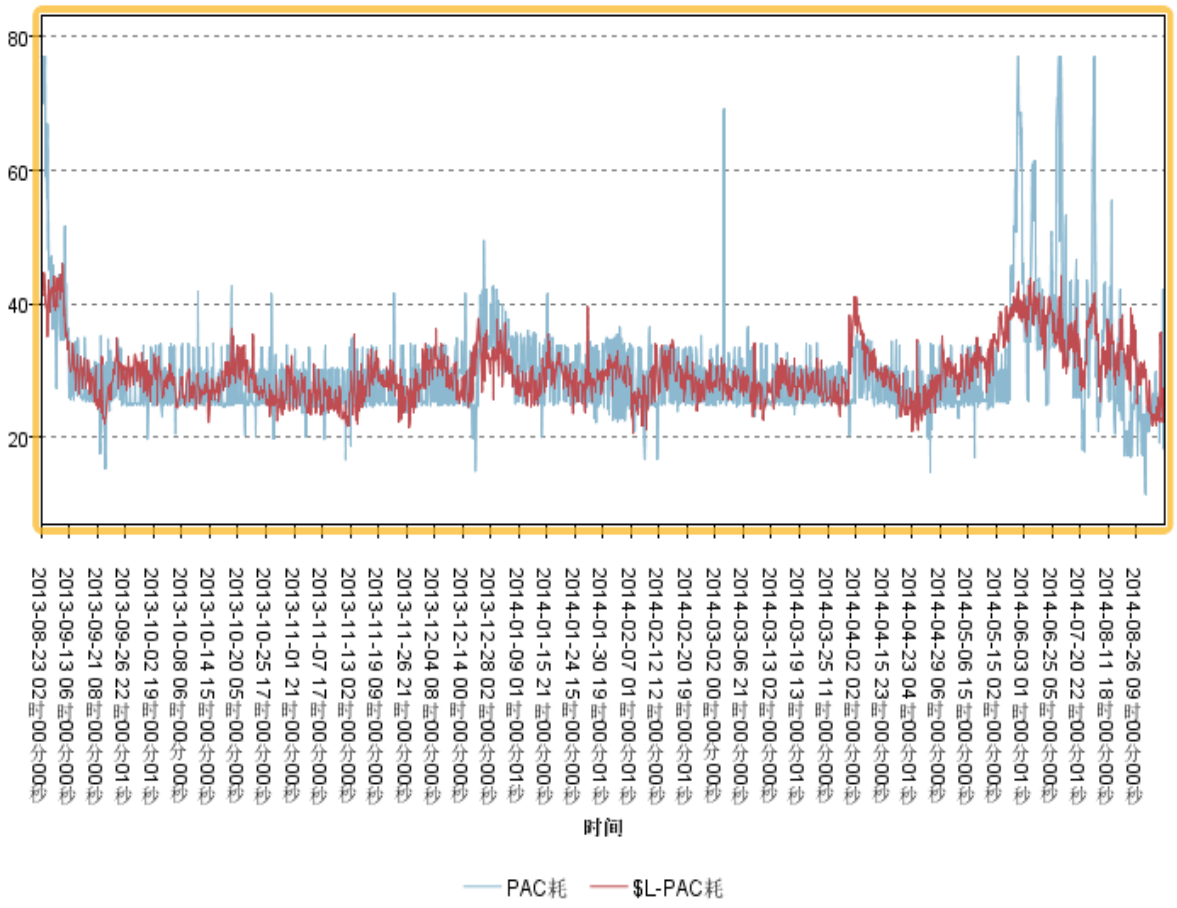


图12. 回归模型训练结果对比

▪2.2.6两种模型的对比分析

表3. BP和多元回归模型性能对比

模型	BP	多元线性回归
相关系数R	0.714	0.526
最大绝对误差	42.049	44.281
平均绝对误差	4	4.574
标准差	5.671	6.75

表3.可以看出，多元线性回归模型的仿真投药量的最大绝对误差、平均绝对误差和标准误差都要比BP网络模型的大，而相关系数比BP神经网络的小。总的来说，BP神经网络的仿真效果要优于回归模型的仿真效果，这是由于混凝投药是一个影响因素多而且不确定的高度非线性系统，在采用多元线性回归建立模型时，我们只能假设它是近似线性的，所以多元线性回归模型并不能很好的反应整个投药系统的各因素之间的关系，而BP神经网络具有高度非线性逼近能力，则能很好的模拟投药系统之间的相互关系，仿真投药量的误差值就比较小。

综上，BP神经网络模型能够很好的对投药量进行仿真，并具有很好的泛化能力，可以根据原水水质等变量对投药量进行很好的预测，明显优于传统的多元线性回归数学模型，而且神经网络还具有一定的学习适应能力，实用性强，可以方便的应用于不同水源的水处理的投药量预测控制。

▪2.2.7 其他因素对最佳投药量的影响。【问题（4）求解】

通常而言，温度也是影响化学反应速度的一个重要因素。这里尝试引入温度数据，分析温度最最佳投药量的影响。原水温度的变化对混凝投药有明显的影响，不同水温下水的动力粘度有很大的差别，水体的表面张力和颗粒的布朗运动也相差很大，混凝的效果也不相同，不同的温度下水的动力黏度见表4.

表4水的动力粘度（$10^{-6}\,pa\bullet s$）

温度℃	0	5	10	15	20	25	30
黏度	1792	1519	1308	1140	1005	897	801

根据文献[8]引入2013年8月至2014年9月的温度数据，并将原始数据按月取平均值，得到数据如附件月份数据表(详见附件3月份数据)。

了解到相关资料表明原水浊度、原水温度、原水PH和投药量之间的关系试表示为[9]：

$$Z_1 = a_0 m_1^{-a_1} Z_0^{a_2} T^{a_3} P^{a_4} \qquad (7)$$

式中，Z_1为沉淀池出水浊度，Z_0为原水浊度，m_1为混凝剂投药量，T为原水温度，P为原水pH，a_1, a_2, a_3, a_4为待估参数，其值均大于零。对上式取对数得到：

$$\ln Z_1 = \ln a_0 - a_1 \ln m_1 + a_2 \ln Z_0 + a_3 \ln T + a_4 \ln P \qquad (8)$$

上式可知取对数之后得到几个变量的线性关系，用表5中的数据取对数后进行回归来确定参数 a_1, a_2, a_3, a_4。得到结果如下表所示：

表5.系数表

系数[a]

模型		非标准化系数		标准系数		
		B	标准 误差	试用版	t	Sig.
1	(常量)	8.999	4.140		2.174	.061
	lnPH	-.599	1.765	-.063	-.339	.743
	ln原水浊度	.197	.096	.526	2.060	.073
	ln原水流量	-.667	.424	-.500	-1.574	.154
	ln出水浊度	.046	.085	.131	.546	.600
	ln温度	.300	.246	.363	1.218	.258

上述表5的系数表中，由于sig 均>0.05 ，即系数不显著，可能由于数据的分布不满足正态性假设，为此，本文又对整体及温度等系数做了正态分布检验，检验结果如表6：

表6 单样本KS检验

单样本 Kolmogorov-Smirnov 检验

		Standardized Residual	ln最佳投药量	ln温度
N		14	14	14
正态参数[a,b]	均值	.0000000	3.4070	3.1305
	标准差	.78446454	.22785	.27596
最极端差别	绝对值	.202	.240	.229
	正	.094	.240	.163
	负	-.202	-.194	-.229
Kolmogorov-Smirnov Z		.755	.899	.857
渐近显著性(双侧)		.619	.394	.454

a. 检验分布为正态分布。

表 6 说明整体与投药量，温度等系数都服从正态分布，但回归分析所得系数确实统计不显著的，由此从统计学角度可以得出温度与最佳投药量线性不相关的结论。所以我们希望从化学角度来分析温度与投药量之间的关系，参考相关文献，我们找到温度对混凝效果不是越高越好，一般情况下原水水温度与絮凝剂所处理的效果有一定关系，原水温度过高或过低对絮凝的作用都是不利的。当水温过高时，化学反应速度过快，使得形成的絮体细小，并使絮凝体的水合作用增加最终产生的沉淀物含水量高，体积大难处理。当水温过低时，有些絮凝剂的水解速度就会减慢，水解时间增长，影响了处理水量。此外水温低时水的粘度增加，增加水对絮体的撕裂作用使絮体变小不利于絮团[10]。

☆2.3 结果分析

由于原始数据中存在很多异常值和缺失值，直接影响后期工作，因此我们运用IBM SPSS MODELER 14.1 用对数据进行了预处理，用 MATLAB 对出水浊度进行三样本插值估算，随后将筛选出来的合格数据用于神经网络模型及多元回归模型的预测。

依据问题的不同要求，我们首先运用平流沉淀理论，求得原水混凝沉淀到出水结束后的时间约为 80 分钟，与实际范围 70min--120min 内相符。

然后确定了 BP 神经网络结构，以原水 PH，原水浊度，原水流量作为 BP 神经网络模型的输入神经元参数，对混凝投药求最佳投药量。为避免输入变量绝对值过大而使得 BP 神经网络的神经元的输出饱和，对数据进行预处理，处理后用于训练神经网络，由误差分析和模型摘要可知，预测值相对稳定，模型训练及预测效果比较好。

为了和 BP 神经网络模型进行比较，运用相同的实验数据建立多元线性回归数学模型，数学模型的参数估计采用最小二乘法计算，利用 SPSS MODELER 软件进行参数求解和模型检验。求出样本回归函数，且模型确立后，用原始建模数据对模型进行仿真效果分析，由残差分布图，并结合仿真投药量和实际投药量的总体曲线走向，可知其仿真效果较好。

本文通过查找水产所在地的气象资料，获取月度温度数据，实证了一个理论模型，但是最后结果表明温度的回归系数的显著性统计性并不显著，表明温度对投药量的影响并不大，从有关化学理论角度对此结果进行解释。

3.结论

随着现代城市的发展，城市供水过程起着越来越重要的作用。目前我国自来水厂常规的治水工艺已经发展得比较成熟了，但制水过程中的自动化水平和管理水平还比较低。因此本文研究了基于 BP 神经网络模型的最佳混凝剂投药量，希望能加速我国水厂自动化的进度。

利用所建立的混凝投药系统神经网络预测模型对某水厂的实际运行数据进行预测，取得了较为满意的结果。特别是随着水厂实际运行的数据不断地增加，并将新的数据添入训练样本数据中，可大大提高网络的学习性能。由此可得，基于神经网络的混凝投药系统预测模型具有较强的自学习性，自适应性，可在此基础上实现前馈神经网络预测控制以实现混凝的最佳投加。

4.参考文献

[1]刘洪波. 平流沉淀池停留时间测定[J]. 城市供水， 2012, (1): 15-7

[2]李壳. 广州南沙水厂的工程设计及特点[J]. 中国给力水排, 2011, (14), 41-45

[3]林洪桦. 剔除异常数据的稳健性处理方法[J]. 中国计量员学报, 2004, (1), 20-24

[4]王金洪. 一维优化三次样条插值法[J]. 哈尔滨师范大学自然科学学报, 2010, (4), 4-8

[5]魏东. 非线性系统神经网络参数预测及控制[M]. 北京:机械工业出版社, 2008, 28

[6]阎平凡, 张长水. 北京: 人工神经网络与模拟进化计算[M]. 2009, 5

[7]田胜元,肖日嵘. 实验设计与数据处理[M]. 北京: 中国建筑工业出版社, 1988

[8]广州市南沙区气候公报[J]. 广州市南沙区气候公报, 2013-2014

[9]杨开明, 张建强, 杨小林. 混凝沉淀过程中最佳混凝剂透亮的研究[J]. 工业水处理, 2005, 25(9): 49-51

[10]赵晓非.表面活性剂对油田污水絮凝效果的影响[J].化学与生物工程,2009,(9):73-75