

基于 Adaboost 算法的水质组合预测方法研究

康 铎¹, 许继平¹, 赵峙尧¹, 王小艺¹, 刘松波²

(1. 北京工商大学 计算机与信息工程学院, 北京 100048; 2. 北京市水务局办公室, 北京 100038)

摘要: 水质预测是水污染防治的重要内容, 针对传统水质预测方法精度低、收敛速度慢等问题, 研究首先选取 Symlets 和 Daubechies 小波系作为小波函数, 对原始数据进行去噪处理并对比, 再结合 RBF、Elman 神经网络以及支持向量机各自优点, 通过不同算法优化 3 种预测模型, 提出基于 Adaboost 算法将优化后的 RBF、Elman 神经网络以及支持向量机相结合的组合预测方法; 以北海为对象进行仿真实验, 验证基于 Adaboost 的溶解氧组合预测方法的有效性, 并分别与单一模型的预测结果进行对比, 结果表明该方法相比于传统的单一模型预测精度得到了提高, 为水质精准预测提供了一种新思路。

关键词: 小波去噪; 水质预测; 神经网络; 支持向量机; Adaboost; 预测器

Research on Water Quality Combination Forecasting Method Based on Adaboost Algorithm

Kang Duo¹, Xu Jiping¹, Zhao Zhiyao¹, Wang Xiaoyi¹, Liu Songbo²

(1. School of Computer and Information Engineering Beijing Technology and Business University, Beijing 100048, China;

2. Office of Beijing Water Authority, Beijing 100038, China)

Abstract: Water quality prediction is an important content of water pollution prevention and control, aiming at the problems such as low accuracy and slow convergence of traditional water quality prediction methods, in this paper, the Symlets and Daubechies wavelet systems are selected as the wavelet function, and the original data are denoised and compared. Combined with the advantages of RBF, Elman neural network and support vector machines, three different prediction models are optimized by different algorithms, a combined forecasting method based on Adaboost algorithm is proposed to optimize RBF, Elman neural network and support vector machine. Taking Beihai as the object to carry on the simulation experiment, the effectiveness of the method based on Adaboost's dissolved oxygen combination is verified and compared with the prediction results of the single model. The results show that the proposed method is improved compared with the traditional single model, which provides a new idea for the accurate prediction of water quality.

Keywords: wave de-noising; RBF neural network; SVM; Adaboost; water quality forecast; predictor

0 引言

随着城市化进程快速推进, 环保基础设施建设未得到同步发展, 城市雨污水在排入时携带较多污染物, 导致了城市水环境形势面临严峻威胁。为了防止水污染扩散, 在河湖关键位置设立了监测站点, 自动监测水质信息状况, 为水污染防治提供信息基础。在此基础上, 根据水质监测信息预测水质变化趋势, 对水环境的有效防范治理具有重要意义。

在水质预测研究中, 数理统计方法和人工神经网络技

术得到了广泛应用。在基于数理统计的水质预测方面, 颜剑波等人建立了多元回归模型^[1], 在自变量和因变量之间规律分析的基础上, 对三门峡断面水质的 COD 浓度进行了预测。荣洁等人应用了指数平滑法—马尔科夫预测模型^[2], 对数据进行平滑处理, 并结合马尔科夫法对合肥湖滨与巢湖裕溪口两大断面的 COD_{Mn}、TP、TN 浓度进行了预测。刘东君等人将最优加权组合预测法应用到永定河的 DO 值的预测^[3]。然而传统的数理统计方法对于数据要求较高, 且影响水质的各因素之间存在着复杂的非线性关系, 使得数理统计方法应用受到限制^[4]。人工神经网络技术目前已成为水质预测的主要研究方向之一。李景文等人利用基于 TSC-RBF 的预测方法, 对漓江阳朔段流域水质中的 COD_{Mn} 和 NH₃-N 进行了预测^[5], 张森等人提出了将偏最小二乘法和支持向量机相耦合的水质预测方法, 以长江朱沱的高锰酸钾指数为例进行分析^[6]。宦娟等人提出了基于 K-means 聚类 and ELM 神经网络的水质预测模型^[7], 应用于养殖水质溶解氧预测。

上述文献中虽然对水质预测方面作了深入的研究, 但并未针对原始数据中的噪声信号进行有效处理, 同时传统的数理统计方法受到多方面的限制, 神经网络方法也只采用了单一预测模型, 不能避免单一方法的局限性。因此本

收稿日期: 2017-12-19; 修回日期: 2018-01-24。

基金项目: 国家水体污染控制与治理重大专项 (2017ZX07104002); 国家自然科学基金 (61703008); 北京市教委科技计划重点项目 (KZ201510011011); 北京市市属高校创新能力提升计划项目 (PXM2014_014213_000033)。

作者简介: 康 铎 (1994-), 男, 北京延庆人, 硕士研究生, 主要从事智慧水务方向的研究。

王小艺 (1975-), 男, 山西运城人, 博士, 教授, 硕士研究生导师, 主要从事污水处理、水华预测方向的研究。

通讯作者: 许继平 (1979-), 男, 湖南岳阳人, 博士, 副教授, 主要从事智能水环境方向的研究。

文将优化后的 RBF、Elman 神经网络以及支持向量机相结合，构建基于 Adaboost 的水质组合预测模型，通过对北海水域中溶解氧浓度进行预测分析，结果表明该模型具有较好的预测精度和泛化能力。

1 基本原理

1.1 小波去噪

小波分析利用小波函数做为基函数，将原始信号按频率分解为多层，各层之间信号不重叠，且所分解的信号包含了原信号的所有频率，其中对于小波函数的选择最为重要^[8]。本文选取 Symlets 和 Daubechies 小波系作为小波函数进行去噪效果对比。

小波去噪的基本步骤如下：

- 1) 结合实际数据，将信号分解为不同频率的信号，并计算每层的小波分解系数。
- 2) 设定各层小波的阈值，阈值处理方式包括软阈值和硬阈值两种。
- 3) 将各层信号与阈值比较并处理后进行小波重构，得到去噪后的信号。

采用的小波去噪效果评价指标为如下两种：

均方根误差 RMSE，表示去噪后信号与原始信号的均方误差，值越小表示去噪效果越好：

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [f(i) - \hat{f}(i)]^2} \quad (1)$$

其中： $f(i)$ 为原始信号， $\hat{f}(i)$ 为去噪后信号。

信噪比 SNR，表示原始信号和噪声值比值，值越大表示去噪效果越好：

$$SNR = 10 \cdot \lg \left(\frac{power_s}{power_n} \right) \quad (2)$$

其中： $power_s$ 表示原始信号功率， $power_n$ 表示噪声功率信号。

1.2 RBF 神经网络

RBF 神经网络是具有单隐层的三层前向网络，其基本思想是，用 RBF 作为隐单元的“基”构成隐藏层空间，隐藏层对输入矢量进行变换，将低维的模式输入数据直接变换到高维空间内，使得在低维空间内的线性不可分问题在高维空间内线性可分^[9-10]。RBF 神经网络的结构示意如图 1 所示。

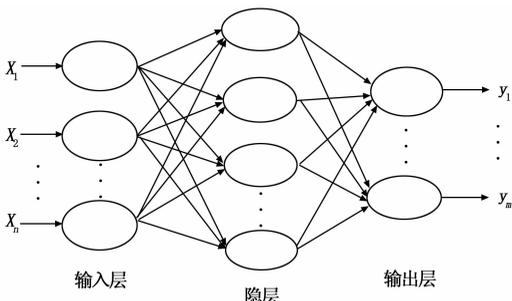


图 1 RBF 神经网络结构图

其中： X_i 为第 i 个输入向量， y_i 为第 i 个输出向量， N 代表隐含层神经元个数。隐含层中的每一个神经元都使用了非线性径向基函数 $\varphi(\cdot)$ ，即输入向量通过输入层到达隐含层，经过隐含层神经元的径向基函数进行非线性变换，再将变换结果加权求和得到输出结果^[11]。

1.3 支持向量机

支持向量机 SVM 可用于模式分类和非线性回归，其主要思想是建立一个分类超平面作为决策曲面，使得正例和反例之间的隔离边缘被最大化。将样本数据从样本空间映射到高维特征空间进行线性回归，从而求解出一个包含了多种因素影响的水质最优回归函数。在最优回归函数中采用适当的核函数代替高维空间中的向量内积，就可以实现非线性变换后的线性拟合，而计算复杂度却没有增加，从而得到最优回归函数^[12]：

$$f(x) = \sum_{i \in SV} (\alpha_i - \alpha_i^*) k(x_i, x) + b \quad (3)$$

其中： α_i, α_i^* 为拉格朗日乘子； b 为回归阈值； SV 为支持向量； $k(x_i, x)$ 为核函数。

1.4 Elman 神经网络

Elman 网络是一种比前向神经网络具有更强计算能力的反馈型神经网络，能够更好的反映系统的动态性^[13]。Elman 网络是在 BP 神经网络基础上多了一个承接层，具有适应时变特性的 Elman 反馈动态递归网络预测性能^[14]。

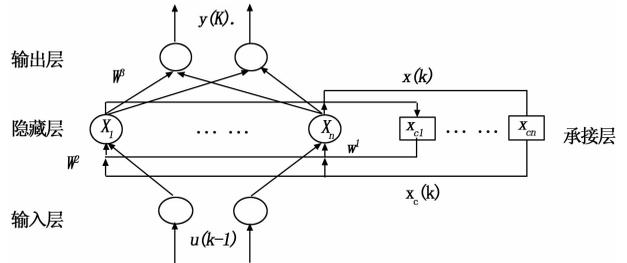


图 2 Elman 神经网络结构图

其中： ω^1 为承接层到隐含层的连接权值， ω^2 为输入层到隐含层的连接权值， ω^3 为隐含层到输出层的连接权值。

1.5 Adaboost 算法

提升 (boosting) 方法是一种常用的统计学习方法，通过改变训练样本的权重，学习多个分类器，并将这些分类器进行线性组合，以提高分类的性能^[15-16]。

Adaboost 方法是最典型的一种算法，其主要思想是：首先给出弱学习算法和 K 组样本数据，每组数据的初始权重为 $1/K$ ；然后利用弱学习算法进行训练，将训练结果与实际值进行比较，将预测失败的训练样本赋予较大权重，使得在下一迭代运算时这些训练样本得到更多的关注，训练失败判定由错误率 ϵ 决定，当 ϵ 大于给定阈值时判定为预测失败样本。从而得到了一系列采用不同权重的训练样本的弱预测器序 f_1, f_2, \dots, f_n ，并且每个弱预测器也具有相应的权值，预测效果越好，弱预测器权值越大；最后在迭代完成后将所有弱预测器加权求和得到强预测器，

再用强预测器进行预测^[17]。

2 组合预测方法

2.1 参数优化

RBF 网络可以自适应的确定网络结构, 但采用固定的目标误差 goal 以及扩展系数 spread, 无法针对样本进行优化, 导致训练精度仍有待提升, 在此使用梯度下降法得出的一系列相对应的目标误差及扩展系数, 通过试凑法获取最优的预测结果。

关于 SVM 参数的优化选取, 并没有公认统一的最优方法, 现在目前常用的方法就是让惩罚系数 c 和核函数半径 g 在一定的范围内取值, 对于取定的 c 和 g 把训练集作为原始数据集, 利用交叉验证方法得到在此组 c 和 g 下训练集验证分类准确率, 最终取使得训练集验证分类准确率最高的那组 c 和 g 做为最佳的参数^[17]。

Elman 网络采取权值更新方法—梯度下降法^[18], 其缺陷是收敛速度慢、容易陷入局部最小值, 在此借助遗传算法 GA 训练初始权值和阈值对 Elman 网络进行优化。改变 Elman 网络依赖梯度下降法来调整网络权值的思想, 利用 GA 全局性搜索的特点, 寻找最为合适的网络连接权值、阈值和网络结构, 提高预测精度和泛化能力。

2.2 组合预测模型建立步骤

以北京北海公园水域作为研究对象, 北海作为地表湖泊且为城市景观水, 其水质指标的变化同时受到自然和人为因素的影响, 对于水环境治理具有指导意义。本文预测模型主要针对北海水域 1~4 月份溶解氧浓度进行分析研究。同时分别采用 RBF 网络、支持向量机以及 Elman 网络进行建模预测, 将预测结果进行对比分析。

第一步, 构造弱预测器。将 3 种单一模型分别作为 3 个弱预测器。在使用神经网络以及支持向量机进行预测时, 本质是找出输入输出之间的非线性函数关系, 要得到良好的预测效果, 输入输出数据的选择很重要。

由于溶解氧浓度的变化是一个渐变的过程, 所以根据历史数据变化趋势来预测后面的输出可以取得较好结果, 如下所示:

$$d(t) = F(d(t-1), d(t-2), \dots, d(t-n)) \quad (4)$$

式中, $d(t)$ 为 t 时刻溶解氧的监测数据, n 为输入层的节点数, F 为由单一预测模型确定的输入—输出映射关系。经试验确定采用最近 3 天的溶解氧历史数据来预测输出效果最好, 即 $n=3$ 。

在模型输入输出选择方面, 分别以溶解氧数据的前 3 天历史数据为模型输入向量, 未来一天的预测值为输出向量, 构建 3 输入 1 输出的弱预测器:

$$d(t) = F(d(t-1), d(t-2), d(t-3)) \quad (5)$$

第二步, 构建 Adaboost 强预测器。用 Adaboost 算法将得到的多个弱预测器序列组成新的强预测器。基于 Adaboost 组合预测方法实现流程如图 2 所示。

Adaboost 组合预测模型具体建模步骤如下:

1) 样本数据权重初始化。首次迭代时数据的初始权重

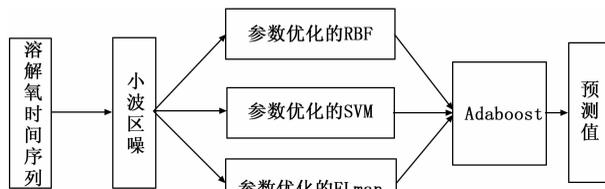


图 3 组合预测方法实现流程图

为 $D_1(k) = 1/n$ ($k=1, 2, \dots, n$ 代表样本序号, 下标 1 代表迭代次数), 为每个样本数据分配相等的权重。

2) 弱预测器预测。每次迭代前将此预测器权值初始化为 0, 再利用相同的训练集训练 3 个弱预测器, 若某一样本数据预测误差大于一定值, 则将其累计权值相加最后得到这一弱预测器的权值之和:

$$Error_j = Error_j + D_i \quad (6)$$

其中: $Error_j$ 代表第 j 个弱预测器权值累加和, D_i 代表超过误差阈值的数据的权值。

3) 更新样本数据权重。若当前样本误差未超过阈值则权值 D_i 不变, 顺延到下一次迭代当中, 若超过误差阈值, 则将权值变大:

$$D_{i+1} = 1.1 \cdot D_i \quad (7)$$

4) 预测序列权值计算。根据弱预测器权值累加和 $Error_j$ 算序列权值:

$$a_{ij} = \frac{0.5}{\exp(|Error_j|)} \quad (8)$$

5) 构建强预测函数。经过 n 轮迭代后得到强预测结果:

$$F = a_i \cdot [f_1, f_1, \dots, f_n] \quad (9)$$

2.3 模型评价指标

选取平均相对误差绝对值 MRE 、最大相对误差绝对值 $MaxRE$ 以及均方误差 MSE 作为对预测结果的评价标准:

$$MRE = \frac{1}{n} \sum_{i=1}^n \left| \frac{X(i) - \hat{X}(i)}{X(i)} \right| \quad (10)$$

$$MaxRE = \max \left(\left| \frac{X(i) - \hat{X}(i)}{X(i)} \right| \right) \quad (11)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n [X(i) - \hat{X}(i)]^2 \quad (12)$$

其中: $X(i)$ 代表真实值, $\hat{X}(i)$ 代表预测值, n 代表样本数量。 MRE 、 $MaxRE$ 和 MSE 的评估标准都为越小越好。

3 仿真实验

3.1 数据预处理

选取 2015 年 1 月至 4 月北海水质监测数据作为训练样本, 2016 年 1 月至 4 月监测数据作为测试数据。选用的水质指标溶解氧浓度, 其中采样频率为每日一次, 两年分别有 120 组数据。选用 Symlets 和 Daubechies 作为小波函数, 阈值处理采用硬阈值, 阈值选择标准为启发式阈值, 分解层数为 5, 分别对原始数据进行去噪处理, 结果如表 1

所示。

表 1 去噪评价指标结果

小波函数	RMSE	SNR	小波函数	RMSE	SNR
db2	0.487	25.024	sym2	0.430	26.088
db3	0.406	26.607	sym3	0.406	26.607
db4	0.335	28.259	sym4	0.356	27.727
db5	0.337	28.206	sym5	0.356	27.727
db6	0.324	28.568	sym6	0.381	27.151
db7	0.343	28.066	sym7	0.355	27.774
db8	0.393	26.871	sym8	0.360	27.643

由表 1 可以看出不同小波函数具有不同的去噪效果,其中 db6 均方误差最低以及信噪比最高,因此采用 db6 作为小波基对数据进行去噪。

3.2 实验及结果分析

将基于 Adaboost 组合预测模型应用于溶解氧浓度预测步骤如下:

第一步,将去噪后的数据归一化处理。数据经过转化映射在 $[0, 1]$ 范围内,采用式 (13) 进行归一化处理:

$$y = (y_{\max} - y_{\min}) \cdot (x - x_{\min}) / (x_{\max} - x_{\min}) + y_{\min} \quad (13)$$

式中, y 为经过处理的数据, x 为原始数据, x_{\min} 、 x_{\max} 分别为最小值和最大值; y_{\max} 和 y_{\min} 分别默认值为 1 和 -1。

第二步,训练和预测。利用所选数据对组合预测模型进行训练和预测,预测精度对比结果见表 2。

表 2 各预测模型预测结果

预测模型	RBF	SVM	Elman	Adaboost 组合预测
MRE	0.005 9	0.010 6	0.049 2	0.004 5
MaxRE	0.053 5	0.084 7	0.142 9	0.024 2
MSE	0.004 1	0.012 2	0.125 2	0.002 0

由表 2 对 2016 年 1~4 月份溶解氧预测精度对比可知,Elman 网络的平均相对误差约在 5%,平均相对误差则更大,达到了 14%以上;RBF 网络和 SVM 相较于 Elman 网络各项指标的预测精度都有了不同程度的提升。而 Adaboost 组合预测模型的预测效果最好,其中均方误差仅有 0.002,相较于前 3 种单一模型,组合模型表现出了明显的优势,预测精度最高,适合于非线性的水质指标预测。单一模型在面对多样环境时,总会有不足,而组合模型恰好的弥补了这一点。

图 3 给出了组合预测模型和 3 种单一模型对于 2016 年 1~4 月份溶解氧预测的误差曲线。

由图 4 可以看出,Adaboost 组合预测模型的误差曲线最低、预测结果最接近真实值,表明了使用 Adaboost 算法对多个不同的弱预测器进行集成,可以有效的提升水质指标的预测精度。

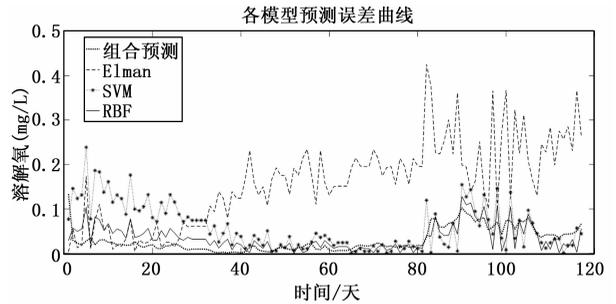


图 4 误差曲线

4 结论

1) 在去噪过程中小波函数的选择对于去噪效果具有显著的影响,在其他条件相同下,db6 小波函数的去噪效果在 Symlets 和 Daubechies 小波系表现最优,有效的去除了数据中的误差信号,为实现水质精准预测奠定了基础。

2) 建立了基于 Adaboost 的组合水质预测模型,通过对北海水域溶解氧浓度历史情况进行分析,并与经过参数优化后的 RBF 网络、支持向量机和 Elman 网络预测结果进行对比,证明相对于传统的单一预测模型,通过 Adaboost 优化的组合预测模型预测结果更加接近真实情况,提高了预测精度和泛化能力。将此方法应用到河湖水质指标变化预测中,可为城市水环境污染防治提供参考依据,满足了人们生产、生活用水的需求。

参考文献:

- [1] 颜剑波, 阮晓红, 孙瀚. 多元回归分析在黄河水质预测中的应用 [J]. 人民黄河, 2010, 32 (3): 35-36.
- [2] 荣洁, 王腊春. 指数平滑法-马尔科夫模型在巢湖水质预测中的应用 [J]. 水资源与水工程学报, 2013, 24 (4): 98-102.
- [3] 刘东君, 邹志红. 最优加权组合预测法在水质预测中的应用研究 [J]. 环境科学学报, 2012, 32 (12): 3128-3132.
- [4] 郑鹏. 改进的 BP 神经网络模型在辽宁中部河流水质预测中的应用研究 [J]. 吉林水利, 2017 (2): 18-21.
- [5] 李景文, 王珂, 殷敏, 等. 基于 TSC-RBF 的水质预测方法——以漓江阳朔河段为例 [J]. 地理与地理信息科学, 2015, 31 (6): 62-66.
- [6] 张森, 石为人, 石欣, 等. 基于偏最小二乘回归和 SVM 的水质预测 [J]. 计算机工程与应用, 2015, 51 (15): 249-254.
- [7] 宦娟, 刘星桥. 基于 K-means 聚类 and ELM 神经网络的养殖水质溶解氧预测 [J]. 农业工程学报, 2016, 32 (17): 174-181.
- [8] 夏巍巍, 万佳威, 丁妍, 等. 基于小波去噪和回归分析的滑坡深部位移预测研究 [J]. 水电能源科学, 2014 (10): 109-112.
- [9] 周开利, 康耀红. 神经网络模型及其 MATLAB 仿真程序设计 [M]. 北京: 清华大学出版社, 2005.

[10] 王小川, 史 峰, 郁 磊, 等. MATLAB 神经网络 43 个案例分析 [M]. 北京: 北京航空航天大学出版社, 2013.

[11] 寇家庆, 张伟伟. 基函数宽度对递归 RBF 神经网络气动力模型精度的影响研究 [J]. 航空工程进展, 2015, 6 (3): 261 - 270.

[12] 梁 坚, 何通能. 基于小波变换和支持向量机的水质预测 [J]. 计算机应用与软件, 2011, 28 (2): 83 - 86.

[13] 李界家, 孙璐璐, 王 奔, 等. 基于 Elman 神经网络阳极效应故障预报方法 [J]. 沈阳建筑大学学报 (自然科学版), 2010, 26 (5): 1012 - 1016.

[14] 关学忠, 宋韬略, 徐延海, 等. 污水处理中 BP 神经网络与

Elman 神经网络的预测比较 [J]. 自动化技术与应用, 2014, 33 (10): 1 - 3.

[15] 李 航. 统计学习方法 [M]. 北京: 清华大学出版社, 2012.

[16] Wang Y Y, Li X. Improvement of RBF neural network by Adaboost algorithm combined with PSO [J]. TELKOMNIKA, 2016, 14 (3): 56 - 64.

[17] 史 峰, 王小川, 郁 磊, 等. MATLAB 神经网络 30 个案例分析 [M]. 北京: 北京航空航天大学出版社, 2013.

[18] Madhiarasan M, Deepa S N. ELMAN Neural Network with Modified Grey Wolf Optimizer for Enhanced Wind Speed Forecasting [J]. 2016, 7 (10): 2975 - 2995.

(上接第 13 页)

面设计上还能更易用直观; 将系统对外开放, 以便其它开发者可以提出新的想法; 进一步改善头盔的集成度与舒适度等。

6 结束语

本文针对重点行业领域安全信息化不足的问题, 设计了一种多场景智能头盔监测系统。着重介绍了系统的架构与方案, 通过功能分层, 阐述了采集终端、数传单元、监测平台的设计与实现。采集终端以 STM32 为核心, 集成在头盔中提高了便携性。环境参数与工作画面根据使用场景的不同, 通过 OpenWrt 路由模块或安卓手机进行无线传输。监测平台也充分考虑了实用性, 提供了 PC 端和移动端两种方式。本系统操作简单且界面友好, 后期也有很好的扩展性, 运行时易于人员维护及分析, 降低了事故发生率。

参考文献:

[1] 刘燕燕, 杨帮华, 丁丽娜, 等. 基于 stm32 的红外火灾探测系统设计 [J]. 计算机测量与控制, 2013, 21 (1): 51 - 53.

[2] 倪天龙. 单总线传感器 DHT11 在温湿度测控中的应用 [J]. 单片机与嵌入式系统应用, 2010 (6): 60 - 62.

[3] 周子昂, 徐 坤, 吴定允. 一种便携式井下多参数监控系统的设计及实现 [N]. 河南理工大学学报 (自然科学版), 2017, 36 (2): 100 - 104.

[4] 刘 练, 周福星. 基于 APP 的智能家居环境监测系统的设计与实现 [J]. 计算机测量与控制, 2014, 22 (7): 2018 - 2023.

[5] 徐大诏, 李正明, 刘 军. 基于 STM32 的便携式矿用多气体检测仪的设计 [J]. 仪表技术与传感器, 2014 (3): 14 - 17.

[6] 徐朝春. 无线网络在机电设备远程监控中的应用 [J]. 能源与节能, 2017 (6): 158 - 159.

[7] 翟伟良, 李 哲. 基于 WiFi 无线传输远程视频监控的研究 [J]. 电子科技, 2016, 26 (9): 67 - 68.

[8] 李丽娟, 李理敏, 郑诺等. 基于 OpenWrt 的无线视频监控系统设计 [J]. 温州大学学报 (自然科学版), 2015, 36 (4): 50 - 55.

[9] 顾富强. 基于 Web 的多点视频交互系统的设计与实现 [D]. 北京: 北京工业大学, 2012.

[10] 潘加玉. 基于无线传感器网络的井下甲烷浓度检测系统 [D]. 沈阳: 沈阳工业大学, 2016.



图 7 PC 端监测界面图



图 8 移动端推流与监测界面图

能够很好地满足用户使用。如果网络状况好, 画面的帧数还能进一步提高。

当然, 本系统的设计上还有些许不足。比如, 软件界