SCIENTIA SINICA Informationis





室内空气质量物联网部署、优化和数据预测

张楠1, 白子轩1, 张泓亮2, 宋令阳1*

1. 北京大学现代通信研究所, 北京 100871, 中国

2. Department of Electrical Engineering, Princeton University, Princeton 08544, USA

* 通信作者. E-mail: lingyang.song@pku.edu.cn

收稿日期: 2021-03-04; 修回日期: 2021-05-19; 接受日期: 2021-06-08; 网络出版日期: 2022-01-10

国家自然科学基金 (批准号: 61931019, 61625101, 61941101) 资助项目

摘要 调查研究显示,城市地区人们平均有超过 70% 的时间在室内度过,因此良好的室内空气质量及 其实时信息对维持健康和生产力至关重要.本文使用低成本物联网传感器网络对室内健身房空气质量 进行监测,研究了电量受限的传感器网络位置部署和能量控制优化问题.为了提高测量数据的有效性, 使用模拟退火算法寻找最优的部署位置;为了提高能量利用率,使用动态规划算法合理规划传感器的 感知策略;为了获得室内空气质量的变化趋势,使用 LSTM 网络对室内空气质量进行预测,并使用遗 传算法对网络超参数进行优化.通过仿真评估所提出的解决方案的性能,显示当联合采用位置部署和 能量优化策略时,性能显著提升,预测算法可以对未来的空气质量进行有效预测.

关键词 空气质量,物联网,位置部署,能量优化,数据预测

1 引言

随着大气污染的日益严重,人们对空气质量越来越关注,有关空气质量的测量得到了较快的发展. 研究报告^[1]显示,空气污染已成为最大的环境污染,在人们关注室外空气质量的同时,室内空气质量 反而容易被忽视,城市中的人们有 70% 的时间在室内度过^[2],室内空气污染对身体健康的影响更为显 著.室内空气污染中,PM2.5 的危害较为严重,主要是指当量直径小于 2.5 μm 的颗粒,其能够长时间 悬浮在空中,携带有毒物质,可以到达肺部,进入肺泡,其携带的有害物质甚至能够进入身体内循环.

随着室外污染的加重以及室外运动场所的受限,人们更倾向于室内运动健身.人在运动时需要消耗大量的氧气,呼吸次数更多,因此室内环境健康对于室内健身运动至关重要,特别是 PM2.5 对人的危害更加严重,主要会引起呼吸系统的疾病^[3],但是室内空气质量变差,仅凭视觉和嗅觉难以发现,因此,实时监测健身房空气质量刻不容缓.然而相对于室外空气质量测量,室内空气质量测量有很大挑战: (1) 室内环境空间小,空气流动性差,空气质量存在区域性分布,需要测量的位置较多; (2) 预测难

引用格式: 张楠, 白子轩, 张泓亮, 等. 室内空气质量物联网部署、优化和数据预测. 中国科学: 信息科学, 2022, 52: 160–175, doi: 10.1360/SSI-2021-0075 Zhang N, Bai Z X, Zhang H L, et al. Deployment, optimization and data prediction of Internet of Things for indoor air quality (in Chinese). Sci Sin Inform, 2022, 52: 160–175, doi: 10.1360/SSI-2021-0075

ⓒ 2022《中国科学》杂志社

度大,室内人员流动性大,空气质量受人的影响因素波动较大,难以应用数值模拟方法进行预测; (3) 预测精度要求较高,室内健身房是健身运动的主要场所,人员活动量大,呼吸频次高,对空气质量分布及预测要求较高.

随着物联网技术的发展,当前室内空气质量测量主要是利用物联网系统进行^[4~12]. 从数据传输方 式来看,可分为有线传输^[4]和无线传输,无线传输中可分为通过 WiFi 网络接入^[5~8]或者利用 ZigBee, Mesh 等无线通信协议组网进行数据传输^[9~11];从节能角度来看,主要有多次测量一次上传^[8],以及将 数据集中后进行上传^[11]. 以往的方法主要是使用有线网络或者 WiFi 网络进行数据的收集和上传,但 是对于大型的室内场所, 网络覆盖并不完善, 传感器的部署会受到网络覆盖的限制, 测量数据不能够 完全体现室内空气质量的分布特性.本文使用 GPRS 模块进行数据上传, GSM 网络覆盖广,传感器的 部署不会受到室内有线网络或者 WiFi 网络覆盖不完全的的限制,可以在任意位置进行部署, 测量数 据能够充分体现室内空气质量的分布特性.

为了对室内测量区域提供全覆盖,物联网空气质量传感设备通常在没有外部电源^[13,14]的情况下 部署,在受限的电池能量约束下,为了延长系统的工作时间,提高系统的能量利用率,需要对传感器的 感知策略进行优化,使传感器进行适当的休眠.目前主要有固定的能量控制和模糊控制^[15],本文使用 动态规划算法对传感器的感知策略进行科学规划.室内空气质量分布不均匀,如何在有限的能量约束 下提高每次测量数据的有效性,使测量数据能够代表空气质量的分布,主要是降低测量数据的相关性, 降低数据冗余度,因此测量设备的位置部署对于描述室内空气质量分布至关重要.以往的部署方法主 要为最大化覆盖面积^[16],没有考虑数据的分布特性,本文使用模拟退火算法对传感器进行优化部署. 为了减少室内空气污染对人身体健康造成的危害,人们需要根据空气质量变化趋势科学安排工作时间 或者提前开启空气净化器对室内空气质量进行改善,因此有必要对变化趋势进行预测.可以使用神经 网络进行预测^[17],本文使用考虑多种室内特征的LSTM 神经网络对空气质量进行预测.

本文主要的研究工作如下: (1) 实际部署系统在校内健身房, 进行为期 2 个月的数据测量; (2) 提 出了一种基于模拟退火算法^[18,19] 的室内传感器部署优化方法, 使测量的室内空气质量分布精度较高; (3) 使用动态规划算法^[20,21] 对传感器网络的感知策略进行优化, 延长传感器网络的工作时间, 提高生成的空气质量分布精度; (4) 根据历史数据, 设计了包含多种输入特征的神经网络对健身房的空气质量 进行预测.

本文其余部分组织如下:第2节对系统整体情况进行介绍;第3节给出了系统模型,并给出了联 合误差的最小化问题;第4节提出了位置部署推荐的模拟退火解决方案;第5节提出了能量控制的动 态规划解决方案;第6节给出了数值预测的 LSTM 网络解决方案;第7节给出了所提解决方案的仿真 结果;最后第8节总结本文.

2 系统概述

本节首先简要概述空气质量测量系统的设计和部署,最后描述收集的数据集.

系统框架如图 1 所示,整个系统可以分为 4 层,分别是感知层、传输层、处理层、展示层.感知 层主要是从环境中获取数据,由部署在室内健身房的 15 台测量设备和温湿度传感器以及 1 台室外测 量设备组成, PM2.5 传感器采用益杉科技的 A3-IG 型激光传感器,微控制器使用 ATmega128A 芯片, 整个系统工作在低功耗模式,采样频率为 5 分钟每次;温湿度传感器采用 DHT11 型号,以树莓派为平 台,间隔 5 分钟进行数据采集,将测量的温湿度数据通过网络写入数据库.传输层主要是将感知层获 取的数据通过 GPRS 模块传输至服务器处理程序,处理层主要是对接收的数据进行提取、校验后存储



图 1 系统框架 Figure 1 System framework

在数据库中,使用动态规划算法对传感器网络的感知策略进行优化,应用神经网络对历史数据和推断数据进行训练,生成预测模型.展示层主要是向用户展示生成的实时空气质量分布图和未来1小时的预测趋势图,如图2所示.

系统在校内健身房部署了 2 个月, 健身房长 32 m, 宽 15 m, 传感器部署在离地面 1.5 m 左右的位置, 由电池进行供电, 当数据传输回服务器, 经过处理后用户可以通过微信公众号查询实时空气质量分布和变化趋势.

在部署过程中,我们收集了超过 10 万个有效值,主要是 PM2.5 浓度.数据集包含 15 台测量设备、1 台温湿度传感器、1 台室外测量设备的测量数据.

3 优化问题

本节首先对室内场景进行描述,然后根据收集数据的统计特性在 3.2 小节中对测量误差和推断误 差分别进行建模,最后在 3.3 小节中给出位置部署和能量优化联合问题建模.

3.1 场景描述

假设在健身房内, 共有 K 个合适的部署位置, N 个测量设备, 其中 N \ll K, 需要将 N 个测量设 备部署在 K 个可能的位置. 用 K_l 和 K_u 分别代表部署位置和未部署位置, 需要满足 $K_l \cup K_u = K$ 以 及 $K_l \cap K_u = \emptyset$. 每个测量设备的传感器和传输模块消耗大部分的电量, 为了节约电量, 延长系统的工



Figure 2 Presentation layer

作时间,这些设备在大部分时间应该保持睡眠模式.控制服务器负责规划所有设备的感知策略,用 0-1 矩阵 $\Phi_{k\times t}$ 表示能量控制策略,其中 t 表示当前时刻, $\Phi_{k\times t}$ 中 1 表示第 k 个位置的设备在 t 时刻执行 感知操作,进行测量数据并上传, 0 表示第 k 个位置的设备在 t 时刻保持休眠,或者此位置未部署传感 器.对于缺失的数据,服务器需要根据测量数据以及时空关系对其进行估计,如图 3 所示.

3.2 测量和推断误差

本小节基于统计数据对测量误差和推断误差进行建模.

系统数据集由图 4(a) 所示的空气质量变化情况收集而来, 通过分析收集到的数据集¹⁾可以发现室 内空气质量存在固定的统计特性:

$$P[\mu_t = y], \quad t \in [0, T], \quad y \in Y, \tag{1}$$

¹⁾ Data collected from indoor gym set from Peking University. https://github.com/sosukesosuke/Indoor_Gym_PM2.5.git.



Figure 3 Space-time model

其中, μt 是平均空气质量, T 是所考虑的时间窗长度, Y 是所有可能的取值空间. 通过对室内空气质量的变化情况进行统计,发现对于恒定的采样频率有

$$P[\mu_t = y | \mu_{t-1} = y'], \quad t \in [0, T], \quad y, y' \in Y.$$
(2)

测量误差. 整个系统主要由传感部分和处理部分组成, 传感部分被部署在健身房需要测量的位置. 令 $X_{k,t}$ ($k \in K_l, t = 0, 1, 2, 3, ...$) 表示第 k 个部署位置 t 时刻的测量值. 传感器的测量值存在一定的测量误差, 这是由传感器本身的特性决定的. 经过长时间的统计分析, 单个设备的测量值相对于同一位置多个测量设备整体均值近似服从高斯 (Gauss) 分布, 如图 4(b) 所示.

$$\begin{cases} X_{k,t} \sim N(\mu_{k,t}, \sigma^2_{k,t}), \\ \mu_{k,t} \approx \mu_t, \\ \sigma^2_{k,t} \approx \mu_{k,t} \times \sigma_0^2, \\ \forall t \ge 0, \ k \in K_l, \end{cases}$$
(3)

其中 $\mu_{k,t}$ 为第 k 个位置第 t 个时刻的精确值, μ_t 为当前时刻的平均值, σ_0^2 为归一化测量方差, 是一个 常数. 由式 (3) 可知, 当空气质量变差时, 测量的结果误差也较大.

时间推断误差. 可以根据第 k 个位置第 t 个时刻的测量值 $X_{k,t}$ 估计 $t+\tau$ 时刻的值. 随着时间的 推移, 新值将随机地偏离原来的值, 这种偏移可以看作是附加的随机噪声施加在原始实测值上, 当时



图 4 历史数据统计特性. (a) PM2.5 时域上的变化曲线; (b) 测量误差; (c) 时间推断偏差; (d) 空间推断偏差 Figure 4 Statistical characteristics of historical data. (a) PM2.5 change curve in time domain; (b) measurement error; (c) time inferred deviation; (d) spatial inferred deviation

隙长度固定时,相邻两个时隙之间的偏差有一个固定的分布,如图 4(c) 所示,满足

$$X_d^{t \to t+1} \sim N(0, \sigma_d^2), \quad \forall t \ge 0, \ k \in K_l,$$

$$\tag{4}$$

其中 σ_d^2 是表示基于给定时间长度的空气质量平均变化率的常数,称为时间偏离方差.

$$X_{k,t+\tau} = X_{k,t} + X_d^{t \to t+1} + \dots + X_d^{t+\tau-1 \to t+\tau},$$
(5)

当相隔 τ 个时间段时有

$$\begin{cases}
X_{k,t+\tau} \sim N(\mu_{k,t}, \sigma_{k,t+\tau}^2), \\
\sigma_{k,t+\tau}^2 = \sigma_{k,t}^2 + \tau \sigma_d^2, \\
\forall t \ge 0, \ \forall k \in K_l.
\end{cases}$$
(6)

这意味着时间跨度越大, 推断的准确性就越低.

空间推断误差. 从历史数据中挖掘不同地点之间的相关性,发现不同地点之间的偏差也可以作为相加的模型,服从高斯分布,如图 4(d) 所示. 因此可以用测量值去推断未部署传感器位置的值,从而形成细粒度的室内空气质量分布.

$$\begin{cases}
X_d^{k \to k'} \sim N(\mu_{k,k',t}, \sigma_{k,k',t}^2), \\
\mu_{k,k',t} \approx \mu_t \times \mu_{k,k'}, \\
\sigma_{k,k',t}^2 \approx \mu_t \times \sigma_{k,k'}^2, \\
\forall t \ge 0 \; \forall k, \; k' \in K,
\end{cases}$$
(7)

其中 $\mu_{k,k'}$ 为使用当前位置推断 k' 位置时平均值的变化量, $\sigma_{k,k'}^2$ 为使用 $X_{k,t}$ 推断 $X_{k',t}$ 增加的归一 化方差, 因此可以推断位置 k' 的值为

$$X_{k',t} = X_{k,t} + X_d^{\ k \to k'},\tag{8}$$

满足高斯分布:

$$\begin{cases}
X_{k',t} \sim N(\mu_{k',t}, \sigma^2_{k',t}), \\
\mu_{k',t} = \mu_{k,t} + \mu_{k,k',t}, \\
\sigma^2_{k',t} = \sigma^2_{k,t} + \sigma^2_{k,k',t}, \\
\forall t \ge 0, \ \forall k, k' \in K.
\end{cases}$$
(9)

由此可知, 当空气质量 µt 变大, 即空气质量变差, 推断误差也就相应增大.

可以进一步利用多个位置的测量值或者经时间推断的估计值来推断未知位置的值,先对每一个目标位置进行单源空间推断,将 M 个推断结果的概率密度函数相乘得到目标位置的概率密度函数,假设每个源相互独立,推断的结果也服从高斯分布,这 M 个推断结果的加权平均值是这 M 个随机变量的平均值的线性叠加,且与单个随机变量相比,方差更小.

$$\begin{cases}
X_{k,t} \sim N(\mu_{k,t}, \sigma_{k,t}^{2}), \\
\mu_{k,t} = \frac{\sum_{m=1}^{M} \mu_{k,t,m} / \sigma_{k,t,m}^{2}}{\sum_{m=1}^{M} 1 / \sigma_{k,t,m}^{2}}, \\
\sigma_{k,t}^{2} = 1 / (\sum_{m=1}^{M} 1 / \sigma_{k,t,m}^{2}), \\
\forall t \ge 0, \quad \forall k \in K.
\end{cases}$$
(10)

3.3 位置部署和能量优化联合问题建模

本小节将对位置部署和能量优化联合问题进行建模.

假设每个测量设备在电池能量耗尽前只能执行 *E* 次感知操作, 且有 *E* < *T*, 即 $\sum_{t=0}^{T} \Phi_{k,t} \leq E$, 为了不让设备长时间休眠, 设置最大休眠时间为 U_{max} , 也就是在 U_{max} 时间段内必须至少感知一次, 即 $\sum_{t=t_1}^{t_1+U_{\text{max}}} \Phi_{k,t} \geq 1$. 为了避免矛盾必须保证 $U_{\text{max}} \cdot E \geq T$.

由于服务器需要提供实时的室内 PM2.5 浓度分布,对于当前时刻未测量地点的不完整数据,应 根据 3.2 小节提到的时间和空间推断使用测量值对其进行估计.对于给定时间段,利用推理方法建立 当前空气质量分布图,为了量化考察该图的准确性,对于 *X*_{k,t},定义联合误差作为数据可靠性的量化 指标:

$$J_{k,t} = \sqrt{\sigma_{k,t}^2 + (\mu_{k,t} - \mu_t)^2}.$$
(11)

平均联合误差综合考虑了方差和偏差,方差和偏差越大联合误差越大,数据的可靠性越低.在整 个系统时间内的平均联合误差定义为

$$\bar{J} = T^{-1} K^{-1} \sum_{t=1}^{T} \sum_{k=1}^{K} J_{k,t}.$$
(12)

以实时室内空气质量分布图的平均联合误差最小化为目标:

$$\min_{K_l} \min_{\{\Phi_{k,t}\}} \bar{J},\tag{13}$$

s.t.
$$\sum_{t=0}^{T} \Phi_{k,t} \leqslant E, \quad \forall t \in [0,T], \quad \forall k \in K_l,$$
(13a)

$$\sum_{t=t_1}^{t_1+U_{\max}} \Phi_{k,t} \ge 1, \quad \forall t \in [0,T], \quad \forall k \in K_l,$$
(13b)

$$\Phi_{k,t} = 0, 1, \quad \forall t \in [0,T], \quad \forall k \in K_l,$$
(13c)

$$\Phi_{k,t} = 0, \quad \forall t \in [0,T], \quad \forall k \in K_u,$$
(13d)

$$|K_l| \leqslant N, \quad |K_l| \in K,\tag{13e}$$

$$K_l \cup K_u = K, \quad K_l \cap K_u = \emptyset.$$
 (13f)

式 (13a)~(13d) 代表能量约束,式 (13e) 和 (13f) 代表位置部署约束.

位置部署与能量控制的联合优化是一个非常困难的问题.因此,在本文接下来的部分中,将问题 (13) 解耦为位置部署优化子问题和能量控制优化子问题,并分别在第 4 和 5 节中给出相应的算法.

4 位置部署推荐

本节主要使用模拟退火算法对给定环境的基于固定能量控制策略的位置部署问题进行求解.

室内场所部署传感器由于场地的限制,不能大量部署传感器.当传感器的数量 N ≪ K (室内感兴趣的位置,需要知道空气质量的位置数量),在确定的能量控制策略下,如何在有限的传感器数量的前提下获得更多感兴趣的位置的空气质量数据,形成高精度的室内分布地图.针对未测量或者未部署传感器的位置,基于上面提到的测量误差、时间推断误差、空间推断误差,可以用传感器的实际测量值或者经时间推断的估计值近似地推断未测量或者未部署设备位置的值.如何部署有限的设备,主要为在感兴趣的位置集中选出固定个数的位置用于部署传感器.选择部署位置的目标是部署位置测量值经过时间和空间推断后形成的室内空气质量分布图的精度最高,即平均联合误差最小.针对每种部署方案,可以用平均联合误差的大小代表部署方案的性能.

解决位置部署优化问题,可以使用模拟退火算法求解最优部署策略,算法的流程图如图 5 所示.

(1) 初始化:初始温度 Γ,使用长度为 K 的 0 – 1 向量表示位置选择情况,其中 1 代表位置被选 中,0 代表位置没有选中,进行均匀初始化,其中 1 的个数为 N,等间隔部署传感器.初始解状态为 w (算法迭代的起点),每个 Γ 值的迭代次数为 L. (2) 对 k = 1, ..., L 做第 (3)~(6) 步. (3) 对当前解 w 中 0,1 进行随机的交换,产生新解 w'. (4) 计算增量 $\Delta f = f(w') - f(w)$,其中 f(w) 为式 (12) 平均联 合误差函数. (5) 若 $\Delta f \leq 0$ 则接受 w' 作为新的当前解,否则以概率 exp $(-\Delta f/\Gamma)$ 接受 w' 作为新的 当前解. (6) Γ 逐渐减少, $\Gamma = 0.99\Gamma$, 然后转第 (2) 步. (7) 如果连续 n 个新解都没有被接受或者温度达 到最低温度 Γ_{\min} ,终止算法. 输出当前解作为最优解,解的平均联合误差最小,即最佳的部署方案.



Figure 5 Flowchart of simulated annealing algorithm

5 能量控制优化

测量设备的能量控制问题可以转化为马尔可夫 (Markov) 决策过程, 进而使用动态规划算法进行 求解. 由于每个测量设备部署位置的空气质量不同, 因此需要对每个部署位置测量设备的感知策略根 据实际的空气质量变化情况进行动态规划, 提高测量效率.

下面对单设备能量优化问题进行建模,建立了一个具有离散和有限状态空间的马尔可夫模型,该 模型描述了能量控制过程中的状态转换.

5.1 马尔可夫过程

在系统转换过程中, 定义第 i 个状态为

$$S_i = (S_i^t, S_i^p, S_i^d, S_i^r, S_i^e).$$

每个状态由 5 个元素构成, $S_i^t \in [0, T+1]$ 为整数, 代表系统时间; $S_i^p \in [0, E]$ 为整数, 代表设备剩余电 量; $S_i^d \in [0, U_{\text{max}}]$ 为整数, 代表设备休眠时间; $S_i^r \in Y$, 记录上一次测量的数值; Y 为 PM2.5 浓度可能 的取值, 为整数; $S_i^e \in Y$, 代表当前 PM2.5 浓度的数值.

初始状态: $S_1 = (1, E, 1, \mu_0, \mu_1)$ 表示设备在第 1 个时刻处于满电状态, 且在零时刻进行了 1 次测量, 当前时刻距离上次检测为 1 个时间间隔.



图 6 状态转移图 Figure 6 State transition diagram

动作集: $A \in \{a_0, a_1\}$, 其中 a_0 表示传感设备休眠, a_1 表示传感设备执行感知操作. 如果 $S_i^p > 0$ 且 $S_i^d = U_{\text{max}}$, 则 $a = a_1$; 如果 $S_i^p = 0$, 则 $a = a_0$, 其他状态下 $a = a_0$ 或 a_1 , 状态转移情况如图 6 所示. 如果在 S_i 状态时执行 a_0 动作, 则代表在 S_i^t 时刻传感器休眠, 该时刻的平均联合误差为 $\sum_{k \in K} J_{k,S_i^t}$,

将平均联合误差的相反数作为 S_i 状态下执行 a_0 动作的回报值,记为

$$R_{S_i}^{a_1} = -\sum_{k \in K} J_{k, S_i^t}.$$
 (14)

如果在 S_i 状态时执行 a_1 动作,则代表在 S_i^t 时刻开启传感器. 该时刻的平均联合误差为 $\sum_{k \in K} J_{k,S_i^t}$,将这个平均联合误差的相反数作为 S_i 状态下执行 a_1 动作的回报值,记为

$$R_{S_i}^{a_1} = -\sum_{k \in K} J_{k, S_i^t}.$$
 (15)

无论是执行动作 a_0 还是 a_1 ,时间元素 S_i^t 在每一次状态转移时都增加 1,当 $S_i^t = T+1$ 时,系统终止.

状态 – 价值函数: 对于任意的系统状态 S,存在一个描述该状态效用的函数,记作 V(S). 当 $S_i^t = T+1$ 时,系统状态终止,终止状态的价值为零,记作 $V(S_{T+1}) = 0$. 中间的任何一步,在状态由 S_i 转移 至 S_{i+1} 时,对应的回报为 $R_{S_i}^a$,则有

$$V(S_i) = R^a_{S_i} + V(S_{i+1}).$$
(16)

由后往前递推,可以得到状态 S₁的价值函数为后续所有回报值的和:

$$V(S_1) = \sum_{i=1}^{T} R_{S_i}^a = -\sum_{t=1}^{T} \sum_{k \in K} J_{k,t}.$$
(17)

因此最大化 $V(S_1)$ 的值 (小于 0) 就是最小化平均联合误差 \overline{J} (大于 0).

5.2 动态规划算法

此问题具有结构化的马尔可夫性,所有状态只与上一个时间点的状态有关,正如文献 [22] 给出的 那样,对于一个确定的马尔可夫过程,始终存在最优的确定性的动作策略,因此可以使用动态规划算 法进行求解.因为最终状态 (能量耗尽)的价值函数为 0,因此可以从最终状态开始,层层反推上一时 刻的所有状态的价值,表示为

$$V(S_i) = \max_{a=a_1,a_0} \left[R_{S_i}^a + \sum_{S_{i+1}} P_{S_{i,S_{i+1}}}^a V(S_{i+1}) \right],$$
(18)

$$\pi(S_i) = \arg \max_{a=a_1,a_0} \left[R_{S_i}^a + \sum_{S_{i+1}} P_{S_i,S_{i+1}}^a V(S_{i+1}) \right].$$
(19)

由式 (18) 可知, 计算 $V(S_i)$ 时, 需要计算所有可能的 S_{i+1} 状态对应的价值函数以及操作奖励.因为 计算 $V(S_i)$ 时考虑了所有可能的后续状态,因此由式 (19) 得出的 $\pi(S_i)$ 是最优的.

依据单设备能量优化方法对每个部署位置的测量设备进行感知策略的动态规划,从而形成整个系统的感知策略,提高系统的测量效率,降低生成的室内空气质量分布图的平均联合误差.

6 数值预测

系统测量设备经过位置部署优化和能量控制优化后,可以生成高精度的室内实时空气质量分布 图,但是未来空气质量的变化趋势同样重要,因此有必要进行预测.首先根据历史数据和时空关系对 未测量数据进行推断,其次使用 LSTM 网络对未来 1 小时的空气质量进行预测,最后使用基因遗传算 法^[23]对 LSTM 网络进行优化,提高预测性能.

6.1 数据推断

通过查询数据库可以获得部署位置当前的测量值,为了获得高精度的室内空气质量分布,需要对 未测量的位置进行数据推断,首先利用测量值应用式 (6) 对未测量值进行时间推断.根据式 (10),利 用全部测量值和时间推断值对所有未部署传感器的位置进行联合空间推断,得到每个推断位置的高斯 分布.

6.2 神经网络预测

室内空气质量的好坏、成因非常复杂,不确定性因素众多,具有非线性特性. 传统的数值模拟方法 计算量大、复杂度高,不能考虑影响空气质量的所有因素,实时性有待提高. 神经网络对于非线性问题 具有很好的拟合性,比较适合对非线性问题进行求解,前提是需要大量的数据进行训练. 对于 PM2.5 测量数值比较容易获取,数据集由实际系统采集而来,大量的历史数据包含了空气质量变化的内在特 征和规律.

特征选择: 未来的空气质量浓度受当前和之前时刻空气质量的影响,因此可以使用当前和之前的 值作为神经网络的输入特征. PM2.5 的测量值对温湿度比较敏感,因此在进行神经网络训练时输入数 据应包含温湿度传感器采集的温湿度数据. 由历史数据可知,室内空气质量和室外空气质量正相关, 室内的 PM2.5 主要是由室外渗透进来的,因此特征中包含室外 PM2.5 测量值.

数据集生成:数据集由室内 15 个测量设备、温湿度传感器、1 个室外测量设备等历史数据构成, 根据网络的输入输出关系形成对应的监督数据.

LSTM 神经网络参数: LSTM 神经网络, 输入数据为 15 个室内测量设备的测量数据和 1 个室 外测量设备以及室内温湿度传感器的数据, 共计 18 个输入参数. 使用之前 z 组数据对未来 1 小时的 数据进行预测, LSTM 单元内的隐藏层的尺寸根据经验设置为 h, 输出层为全连接层, 由于需要预测未 来 1 小时的数据, 采集数据时间间隔为 5 min, 因此输出层单元数为 12.

6.3 神经网络优化

神经网络的超参数一般主要靠经验进行设置,本文中利用遗传算法对神经网络的超参数 z (输入数据的第 2 个维度即时间步) 和 h (LSTM 单元内的隐藏层的尺寸) 进行优化,达到提高预测精度的目的.

基因编码和初始化: 对网络超参数 *z* 和 *h* 进行二进制编码, 每个参数使用 6 个比特进行编码, 随 机生成 *N*₀ 个基因, 对种群进行初始化.

基因选择:首先计算种群中每个基因的适应性,这里使用基因对应的超参数设置的网络训练后在测试集上的预测均方根误差 (root mean squared error, RMSE) 为评价标准,按照从小到大对基因性能进行排序,选择其中前 N₁ 个基因,从剩下的基因中随机选择 N₂ 个基因.

基因重组: 基因重组是产生潜在解的主要方式,任意两个基因在随机位置交换之后生成新的基因, 共产生 N_3 个,其中 $N_0 = N_1 + N_2 + N_3$.

基因突变: 基因突变是产生潜在新解的一种方式,每个比特以 P 的概率从 0 变为 1 或者从 1 变为 0.

经过 n 次进化后选出其中最优秀的个体,也就是 RMSE 最小的的个体,将选出的个体进行解码作为神经网络的最优参数用于数据预测.

7 仿真结果和讨论

7.1 位置部署推荐

在模拟退火算法仿真中参数 K = 15, n = 10, L = 100, T = 1000, t = 0.01, N = 5, 6, 7, 8, 9, 10. 通过 仿真,随着测量设备的增加,两种部署方案的平均联合误差都减小,在设备数量从 5 增大到 10 的过程 中优化部署方案的平均联合误差均小于随机部署方案的平均联合误差,如图 7 所示.优化部署方案可 以获得精度更高的室内空气质量分布,但是随着设备数量的增加,两种部署方案的平均联合误差差距 减小,在 N = 10 的情况下,优化部署相对于随机部署可以获得 25.1% 的性能增益.

7.2 单设备能量优化

仿真中 T = 500, E = 100, U_{max} = 12,使用动态规划算法获得的感知策略如图 8 中短线所示,可以 看到,在 PM2.5 浓度变化剧烈的位置短线比较密集,在变化比较缓慢的时候短线比较稀疏,表示 PM2.5 浓度变化剧烈的时候设备测量频率加快以降低误差, PM2.5 浓度变化缓慢的时候设备测量频率减慢以 节约电量,在相同参数的情况下动态规划算法获得的平均联合误差为 2.39,均匀感知策略获得的平均 联合误差为 2.71,动态规划算法相较于均匀感知可以获得 15.1% 的性能增益.

7.3 多设备能量优化

仿真中,参数 T = 500, E = 100, U_{max} = 12, K = 15, N = 5, 6, 7, 8, 9, 10. 如图 9 所示,在优化部署的位置选择策略下,能量控制分别使用随机感知策略、均匀感知策略、动态感知策略进行仿真对比.可以看到随着设备数量的增加, 3 种策略的平均联合误差均逐渐减小,动态感知策略性能最优,均匀



图 7 优化位置部署和随机位置部署性能对比 Figure 7 Optimize location versus random deployment



图 9 多设备能量控制性能对比 Figure 9 Comparison of energy control performance of multiple devices



图 8 单设备动态规划算法能量控制优化 Figure 8 Energy control optimization of single device dynamic programming algorithm



图 10 位置优化和能量控制优化联合效果对比 Figure 10 Combined effect of position optimization and energy control optimization

感知策略次之,随机感知策略最差.在 N = 10 的情况下,动态感知策略较均匀感知和随机感知分别 有 25.9% 和 32.46% 的性能增益.

7.4 位置部署和能量控制联合优化

使用优化的位置部署和动态感知的能量优化策略对系统进行联合优化,其中 T = 500, E = 100, $U_{\text{max}} = 12, K = 15, N = 5, 6, 7, 8, 9, 10.$ 如图 10 所示,可以发现动态感知策略的性能好于均匀感知策略的性能,优化部署策略性能明显好于随机部署策略性能.在 N = 10的情况下,动态感知策略联合优化部署策略相对于随机部署下的均匀感知策略可以获得 44.8% 的性能提升.

7.5 神经网络预测及优化

仿真参数: 输入数据维度为 (19596, *z*, 18), 参数 *z* 和 *h* 根据经验设置为 *z* = 120, *h* = 32, 输出数 据维度为 (19596, 12, 1) 预测未来 1 小时的数据. 利用系统收集的数据和推断数据对神经网络进行训 练, 得到预测模型. 网络模型的预测结果如图 11(a) 所示, RMSE = 1.73.

遗传算法优化仿真参数为 N₀ = 50, P = 0.01, N₁ = 10, N₂ = 20, N₃ = 20, n = 20, 使用基因遗传



Figure 11 The prediction results before and after neural network optimization

算法对神经网络的超参数进行全局寻优,获得的最优参数 z = 47, h = 59,使用最优的参数对神经网络进行训练,获得的预测性能有了进一步提升, RMSE = 1.43 可以获得 17.3% 的性能增益,如图 11(b) 所示.

8 总结

本文主要使用物联网传感器网络对室内健身房空气质量进行监测和展示. 针对室内空气分布不均 匀的特点,设计了基于模拟退火算法的传感器部署优化算法,使传感器的部署达到最优. 针对传感器 电池能量受限的特点,使用动态规划算法对传感器的感知策略进行优化,提高生成的室内空气质量分 布图的精度.设计了一种基于遗传算法优化的考虑多种室内特征的 LSTM 网络,对室内空气质量进行 预测,并将实时的空气质量分布和未来1小时的预测结果在开发的微信公众号上进行展示. 仿真结果 表明,位置推荐部署优化和能量控制联合优化相对于随机部署策略下均匀能量控制,可以有效提高生 成的室内空气质量分布精度,预测算法可以对未来1小时的空气质量进行有效预测.

参考文献

- 1 Launay F. 7 million premature deaths annually linked to air pollution. Central European J Public Health, 2014, 22: 53–59
- 2 Brasche S, Bischof W. Daily time spent indoors in German homes-baseline data for the assessment of indoor exposure of German occupants. Int J Hygiene Environ Health, 2005, 208: 247–253
- 3 Pope C A, Burnett R T, Thun M J, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. J Am Medical Assoc, 2002, 287: 1132–1141
- 4 Chen X X, Zheng Y, Chen Y B, et al. Indoor air quality monitoring system for smart buildings. In: Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing, New York, 2014. 471–475
- 5 Ladekar V, Daruwala R. Indoor air quality monitoring on AWS using MQTT protocol. In: Proceedings of the 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019. 1–6
- 6 Giris S V, Prakash R, Ganesh A B. Real-time remote monitoring of indoor air quality using Internet of Things (IoT) and GSM connectivity. In: Proceedings of Artificial Intelligence and Evolutionary Computations in Engineering Systems, New Delhi: Springer India, 2016. 527–533

- 7 Lohani D, Acharya D. SmartVent: a context aware IoT system to measure indoor air quality and ventilation rate.
 In: Proceedings of IEEE International Conference on Mobile Data Management, Porto, 2016. 64–69
- 8 Yang X K, Yang L Y, Zhang J. A WiFi-enabled in-door air quality monitoring and control system: the design and control experiments. In: Proceedings of IEEE International Conference on Control & Automation, Ohrid, 2017
- 9 Jelicic V, Magno M, Brunelli D, et al. Context-adaptive multimodal wireless sensor network for energy-efficient gas monitoring. IEEE Sens J, 2013, 13: 328–338
- 10 Abraham S, Li X R. A cost-effective wireless sensor network system for indoor air quality monitoring applications. Proceedia Comput Sci, 2014, 34: 165–171
- 11 Kim J Y, Chu C H, Shin S M. ISSAQ: an integrated sensing systems for real-time indoor air quality monitoring. IEEE Sens J, 2014, 14: 4230–4244
- 12 Jha R K. Air quality sensing and reporting system using IoT. In: Proceedings of the 2nd International Conference on Inventive Research in Computing Applications (ICIRCA), 2020. 790–793
- 13 Gubbi J, Buyya R, Marusic S, et al. Internet of Things (IoT): a vision, architectural elements, and future directions. Future Generation Comput Syst, 2013, 29: 1645–1660
- 14 Fioccola G B, Sommese R, Tufano I, et al. Polluino: an efficient cloud-based management of IoT devices for air quality monitoring. In: Proceedings of IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better to-morrow (RTSI), 2016. 1–6
- 15 Pradityo F, Surantha N. Indoor air quality monitoring and controlling system based on IoT and fuzzy logic. In: Proceedings of the 7th International Conference on Information and Communication Technology (ICoICT), 2019. 1–6
- 16 Farsi M, Elhosseini M A, Badawy M, et al. Deployment techniques in wireless sensor networks, coverage and connectivity: a survey. IEEE Access, 2019, 7: 28940–28954
- 17 Jiao Y, Wang Z, Zhang Y. Prediction of air quality index based on LSTM. In: Proceedings of IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 2019. 17–20
- 18 Yao X, Chen G L. Simulated annealing algorithm and its application. Comput Res Dev, 1990, 7: 1-6 [姚新, 陈国良. 模拟退火算法及其应用. 计算机研究与发展, 1990, 7: 1-6]
- 19 Steinbrunn M, Moerkotte G, Kemper A. Heuristic and randomized optimization for the join ordering problem. VLDB J Int J Very Large Data Bases, 1997, 6: 191–208
- 20 Puterman M L. Markov Decision Processes: Discrete Stochastic Dynamic Programming. New York: John Wiley & Sons, 2014
- 21 Eddy S R. What is dynamic programming? Nat Biotechnol, 2004, 22: 909-910
- 22 Sutton R S, Barto A G. Reinforcement Learning: An Introduction. Cambridge: MIT Press, 1998
- 23 Mirjalili S. Genetic algorithm. In: Evolutionary Algorithms and Neural Networks. Cham: Springer, 2019. 43-55

Deployment, optimization and data prediction of Internet of Things for indoor air quality

Nan ZHANG¹, Zixuan BAI¹, Hongliang ZHANG² & Lingyang SONG^{1*}

1. Institute of Modern Communication, Peking University, Beijing 100871, China;

2. Department of Electrical Engineering, Princeton University, Princeton 08544, USA

* Corresponding author. E-mail: lingyang.song@pku.edu.cn

Abstract Research shows that people in urban areas spend on average more than 70% of their time indoors, so good indoor air quality and its real-time information are critical to maintaining health and productivity. In this paper, the low-cost Internet of Things sensor network is used to monitor the air quality of indoor gym, and the position deployment and energy control optimization of the sensor with limited power are studied. In order to improve the effectiveness of the measurement data, the simulated annealing algorithm is used to find the optimal deployment location. In order to improve the energy utilization rate and prolong the working time of the system, the dynamic programming algorithm is used to reasonably plan the sensing strategy of the sensor. In order to obtain the variation trend of indoor air quality, the LSTM network is used to predict the indoor air quality, and the genetic algorithm is used to optimize the network super parameters. The performance of the proposed solution is evaluated by simulation, and it is shown that there is a significant performance gain when the location deployment and energy optimization strategies are adopted, and the prediction algorithm can effectively predict the future air quality.

Keywords air quality, Internet of Things, location deployment, energy optimization, data prediction