文章编号:1671-4598(2021)06-0093-05 DOI:10.16526/j.cnki.11—4762/tp.2021.06.020 中图分类号:V412.4;V448.2 文献标识码:A

基于 Deep Q Networks 的交通指示灯控制方法

颜文胜[1]，吕红兵[2]
(1. 台州职业技术学院 信息技术工程学院，浙江 台州 318000；
2. 浙江大学 计算机科学与技术学院，杭州 310027)

摘要：交通指示灯的智能控制是当前智能交通研究中的热点问题；为更加及时有效地自适应动态交通，进一步提升街道路口车流效率，提出了一种基于 Deep Q Networks 的道路指示灯控制方法；该方法基于道路指示灯控制问题描述，以状态、行动和奖励三要素构建道路指示灯控制的强化学习模型，提出基于 Deep Q Networks 的道路指示控制方法流程；为检验方法的有效性，以浙江省台州市市府大道与东环大道交叉路口交通数据在 SUMO 中进行方法比对与仿真实验；实验结果表明，基于 Deep Q Networks 的交通指示灯控制方法在交通指示等的控制与调度中具有更高的效率和自主性，更有利于改善路口车流的吞吐量，对道路路口车流的驻留时延、队列长度和等待时间等方面的优化具有更好的性能。

关键词：道路指示灯；Deep Q Networks；智能交通；信号控制

Road Signal Light Control Method Based on Deep Q Networks

Yan Wensheng[1], Lv Hongbing[2]
(1. School of Information Technology Engineering, Taizhou Vocational and Technical College, Taizhou 318000, China;
2. College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China)

Abstract: The intelligent control of traffic lights is a hot issue in the research of intelligent traffic. In order to adapt to dynamic traffic in a more timely and effective manner and further improve traffic flow efficiency at street intersections, a road indicator light control method based on Deep Q Networks was proposed. This method is based on the description of the road indicator control problem, constructs the reinforcement learning model of the road indicator control with the three elements of state, action and reward, and proposes the road indicator control method flow based on Deep Q Networks. To test the effectiveness of the method, the traffic data at the intersection of Shifu Avenue and Donghuan Avenue in Taizhou City, Zhejiang Province were compared and simulated in SUMO (simulation of urban MObility). The experimental results show that the traffic light control method based on Deep Q Networks has higher efficiency and autonomy in the control and scheduling of traffic indications, is more conducive to improving the throughput of intersections traffic flow, and has better performance in optimizing the stay delay, queue length and waiting time of intersections traffic flow.

Keywords: road signal lamp; Deep Q networks; intelligent transportation system; signal control

0 引言

现有的道路交叉口管理都是通过红绿指示灯来实现的。固定时间、基于周期的道路指示灯控制模式效率低下，导致车流滞留时间长、能源浪费大、空气质量恶化等问题。在某些情况下，它还可能导致交通事故[1-2]。现有的道路指示灯控制要么以固定的程序，要么没能考虑实时交通[3]，这在应对各种实际情况时，特别是在早晚高峰车流量徒增的情况中将面临交通临时瘫痪的现象。

为进一步改善这一现状，不少学者进行了深入研究。例如文献［4］为提高智能交通信号在实际应用中的效率，提出一种改进的控制方法，将交通信号图像输入卷积神经网络的输入层，通过卷积层与采样层的卷积计算、残差计算以及梯度计算识别交通信号，将识别交通信号结果选取自适应跳跃式信号控制方法实现了智能交通信号控制。文献［5］针对城市单交叉口的交通信号控制问题，基于四相位定相序对单交叉口交通灯进行控制，运用模糊控制系统输入为车辆排队数和车辆到达率，输出为当前绿灯相位的绿灯延长时间，提出一种交通灯信号的模糊控制方法，从而能有效减少车辆的平均延误时间，提高交叉口的通行能力。文献［6］针对城市交通子区内部与边界交叉口协调控制问题，提出基于分层多粒度与宏观基本图交通信号控制模型（HDMF），从而能够有效疏导子区域内部与边界交

收稿日期:2020-11-10； **修回日期**:2020-12-04。
基金项目:浙江省高等教育“十三五”教学改革研究项目(jg20190884)；浙江省教育厅科研项目(Y202044737)；台州职业技术学院 2019 年重大专项课题(2019HGZ02)。
作者简介:颜文胜(1972-)，男，浙江台州人，硕士，副教授，主要从事人工智能的相关方向的研究。
引用格式:颜文胜，吕红兵.基于 Deep Q Networks 的交通指示灯控制方法[J].计算机测量与控制，2021，29(6)：93-97.

通，实现整体路网的车流量最大化。文献［7］在城市交通环境下，通过分析控制方法灵活性与稳定性的关系，将稳定状态引入信号控制决策模块，建立稳定规则库，提出了一种考虑网络稳定性的多智能体强化学习控制方法，从而提高了算法的运行效率，同时保证了控制效果，将适用于复杂交通网络。

随着计算机、通信和交通检测技术的变革式发展，城市道路指示控制系统的技术环境正从数据贫乏向数据丰富的时代演化发展[8-10]。本文发挥 Deep Q Networks 无需固定控制规则、无需同时获取大量数据，而通过不断地从环境中获取状态和奖励进行更新的特性，提出了一种基于 Deep Q Networks 的交通指示灯控制方法。

1 交通指示灯控制的问题描述

在道路交通信号控制场景中，信号灯被用来管理道路十字路口的交通流[11]。道路十字路口的信号灯设有 3 个状态信号：红、黄、绿。通常，道路十字路口会有来自多个方向的车辆涌入，存在有的方向车流量大、有的方向车流量小的情况。当前，信号灯所采用的固定规则模式，难以应对十字路口车流不均的情况，导致十字路口有的方向严重拥堵现象。这一问题在北京、上海等大城市中非常严重，目前采用的是高峰时间段由人工控制的方式进行缓解。

为了能够让车流交替通过十字路口的同时，使得各路口等待车辆数达到均衡，需要调整路口信号灯的持续时间，以应对道路十字路口的不同交通情况。为此，需要解决的问题是如何通过借鉴历史经验，通过动态改变道路指示灯的每个状态的持续时间，以缓解道路十字路口车辆拥堵问题。面对这一问题，通常的想法是延长拥堵方向上绿灯的时间，让该道路上更多的车辆通行。但是，根据当前复杂交通状况，给出一种适用的控制规则是困难的，更难以适应每天状态各异的道路交通。

一种常用的解决方式是韦伯斯特（Webster）法[12]。该方法是一种自适应韦氏交通信号控制算法，通过收集各时段的数据，然后利用韦氏方法计算下一时段的周期和信号灯持续时间。这种自适应 Webster 方法本质上使用最近的时间区间来收集数据，并假设下一区间的流量需求大致相同。其时间区间的选择至关重要，并体现了各种各样的权衡，较小的区间允许更频繁地适应变化的交通需求，而较大的区间不太频繁地适应，但有利于增加稳定性。

$$C_m = \frac{L}{1-\sum_{i=1}^{n} y_i} = \frac{L}{1-Y} \tag{1}$$

式中，C_m 是完成一次车辆通行的时间区间；y_i 是第 i 个相位上交通量最大的车道的车流比；L 是一个信号周期的总损失时间。

$$L = \sum_{i=1}^{n} (l_i - I_i - A_i) \tag{2}$$

式中，l 是车辆启动的损失时间；I 是信号灯为绿的时间间隔；A 是信号灯为黄的时间间隔。

令表示路口车流延误最低的最佳周期时长：

$$C_w = \frac{1.5L+5}{1-Y} \tag{3}$$

Deep Q Networks，则是一种机器学习范式，在这种范式中，个体通过与环境的反复交互，寻求通过制定一种状态－行动政策来最大化累积回报。Deep Q Networks 通过制定最优状态－行动政策来实现奖励的最优控制。Deep Q Networks 是尝试解决道路指示控制问题的一种合适技术，能够通过强化学习三要素对问题进行很好地描述：agent（道路指示控制器）、environment（交通状态）以及 actions（交通信号）。

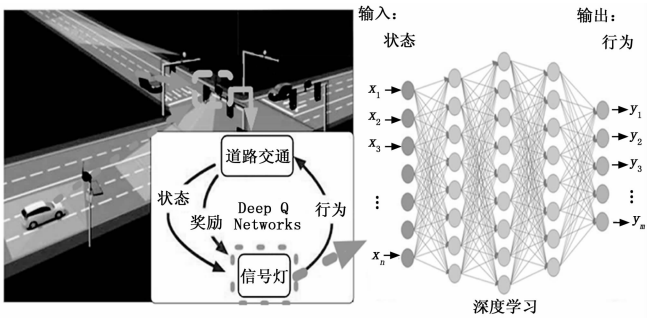

图 1 基于 Deep Q Networks 的交通指示灯控制

基于 Deep Q Networks 的交通指示灯控制如图 1 所示，其中左图是道路指示灯控制示意图。图中信号灯首先通过车载网络[13]采集道路交通信息。信号灯对数据进行处理，得到道路交通的状态和奖励[14]。信号灯使用右图显示的深度学习网络选择下一步动作。信号灯整个这一“强化学习＋深度学习”的自动控制过程，构成了道路交通信号灯控制的 Deep Q Networks 模型。

2 道路指示灯控制的 Deep Q Networks 模型

根据道路十字路口车辆的位置和速度这两个信息定义道路指示灯控制的状态模型。通过车载网络和定位系统，可以很容易地获得当前道路车辆的位置和速度[13]。然后，道路十字路口的信号灯可以通过车辆信息位置矩阵得到当前路口的虚拟快照图像。将当前路口的虚拟快照划分为相同大小的小正方形网格。其中，网格长度 c 应确保只能容纳一辆车。在每个网格中，状态值由位置 p 跟速度 v 组成，即 $[p,v]$。其中，位置 p 是一个二进制值，表示网格中是否有车辆。如果网格中有辆车，则该网格中的值为 1；否则，则为 0。速度 v 为整数值，表示车辆当前速度，单位为 m/s。

图 2 为道路十字路口虚拟快照示意图，整个道路场景被分割为正方形网络。相应的位置矩阵与网格大小相同，如图 3 所示。其中，一个网络单元代表一个位置状态，空白单元表示对应网格中没有车辆，其值设为 0。

道路指示灯需要根据当前的交通状态，选择合适的行

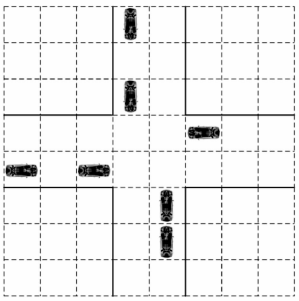

图 2　交通路口车流通行示意图

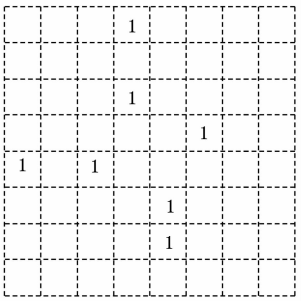

图 3　交通路口车流位置矩阵

为引导路口车辆通行。通常，在道路指示灯控制系统中，通过选择下一周期中红绿灯每一阶段的持续时间来确定行动空间。但是如果相邻两个周期内的持续时间变化很大，系统将可能会变得不稳定。

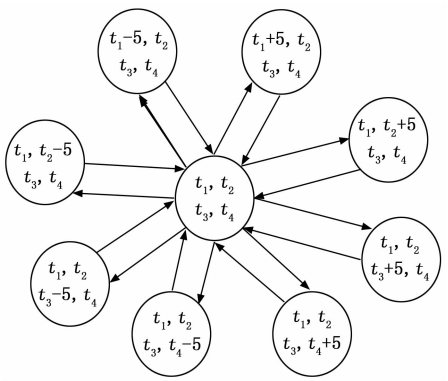

图 4　多指示灯下的 MDP 决策

因此，为了应对道路指示的动态调度，将两个相邻周期之间的持续时间调度建模为一个高维马尔可夫决策(MDP)。MDP 是一个灵活的模型，它可以应用于交通灯较多、需要更多状态的复杂交叉路口，甚至能够满足具有五、六条道路的不规则交叉路口。在 MDP 中，交通灯在一个小步骤中只改变一个阶段的持续时间，本文将采用一个四元组 $[t_1,t_2,t_3,t_4]$ 表示当前周期中 4 个阶段的持续时间。下一个时间周期的行为如图 4 所示。图中，一个圆表示在一个时间周期中 4 个阶段的持续时间，并将当前周期到后续时间变化离散为 5 s。将各阶段的最大合法持续时间设置为 60 s，最小合法持续时间设置为 0 s。道路指示灯根据当前状态选择行为，信号灯的状态按顺序循环变化。

为了保证安全，相邻状态之间需要有黄色信号，使行驶的车辆在信号变为红色之前停止。将黄色信号持续时间 T_{yellow} 的定义为该道路上的最大速度 $v_{\max}$ 除以路口减速率 a_{dec}：

$$T_{\text{yellow}} = \frac{v_{\max}}{a_{\text{dec}}} \tag{4}$$

奖励是区别强化学习与其他学习算法的一个重要特征。奖励的作用是就先前行为的表现向强化学习模型提供反馈。因此，明确奖励方式以正确指导模型自主学习是很重要的，将有助于选择最佳的行动策略。

在道路指示控制系统中，主要目标是要提高十字路口车辆通行效率。衡量车辆通行效率的一个主要指标是十字路口车辆的等待时间。因此，将强化学习的奖励定义为相邻两个周期之间累积等待时间的变化，用 i_t 表示观察的第 i 量车从第 t 个周期开始的时间，用 N_t 表示到第 t 个周期对应的车辆总数，将车辆 i 在第 t 周期的等待时间记为 $\omega_{i_t,t}(1 \leqslant i_t \leqslant N_t)$，则第 t 周期的奖励定义为：

$$r_t = W_t - W_{t+1} \tag{5}$$

其中：

$$W_t = \sum_{i_t=1}^{N_t} \omega_{i_t,t} \tag{6}$$

由式 (5) 可知，如果奖励相较之前有所增加，则等待时间的增量将小于之前。这意味着奖励时采取行动前和行动后累计等待时间的增量。

3　基于 Deep Q Networks 的交通指示灯控制

在交通指示灯控制问题描述的基础上，基于所构建的道路指示灯控制 Deep Q Networks 模型，提出了基于 Deep Q Networks 的道路指示控制方法。

首先，令 $Q(s,a)$ 表示在状态 $s=(X;Y)$ 当采取行动 a 时的行为价值函数，即最大可实现的预期折扣奖励：

$$Q(s,a) = r(s,a) + \gamma \sum_{s' \in S} p(s,s';a) \max_{a' \in A} Q(s',a') = r(s,a) + \gamma E[\max_{a' \in A} Q(s',a')] \tag{7}$$

式中，$p(s,s';a)$ 表示采取行为 a 时由状态 s 到状态 s' 的转移概率；行为价值函数的预测值 $V(s) = \max_{a' \in A} Q(s',a')$ 满足 Bellman 优化方程[1]：

$$V(s) = \max_{a' \in A} \{r(s,a) + \gamma \sum_{s' \in S} p(s,s';a) V(s')\} = \max_{a' \in A} \{r(s,a) + \gamma E[V(s')]\} \tag{8}$$

接着，在强化学习算法中，将状态 s 同时作为目标网络和评估网络的输入，依据式 (8)，构建神经网络迭代 i 次后的训练损失函数模型：

$$L(\theta_i) = E_{s,a,r,s'}[(r + \gamma max Q(s',a';\theta'_i)) - Q(s',a';\theta_i)] \tag{9}$$

式中，r 为当前步骤采取行为所获得的奖励值，s' 和 a' 分别为下一步的状态和行为。

根据式 (9) 在神经网络中进行反向传播并更新主神经

网络中的参数：

$$\theta_i' = \alpha\theta_i' + (1-\alpha)\theta_i \tag{10}$$

式中，α 为更新速率，它表示最新参数对目标网络中各组件的影响程度。

最后，给出基于 Deep Q Networks 道路指示灯控制方法的训练流程：

步骤 1 初始化车流状态 s 和控制行为 a；

步骤 2 对于训练步长 $k=1,2,\cdots,K$；

步骤 3：根据行为价值函数选择行动 $a^* = \arg\max_{a\in A} Q(s,a)$；

步骤 4：给定行动，确定新状态 s'；

步骤 5：根据式（5）、式（6）计算奖励值 r；

步骤 6：将状态、行动和奖励以 $[s,a^*,r,s']$ 形式在记忆库 M 中存储；

步骤 7：判断：如果 $k>k_0$ 执行学习操作；

步骤 8：从记忆库中取样一小批量样本；

步骤 9：根据式（10）更新目标网络及参数 θ'；

步骤 10：根据式（7）计算行为价值函数值 $Q(s,a)$；

步骤 11：运用梯度下降更新网络，利用贪婪策略依式（9）计算损失函数。

4 算例仿真

为了更好地验证本文方法的有效性以及对比优势，在本节分别与固定时间控制模式、Webster 控制模式进行了对比仿真实验。本仿真基于 Intel Core i5cpu 硬件环境，运用微观交通仿真平台 SUMO v0.32 构造道路指示灯仿真场景，实现道路指示灯的自主控制。方法模型运用 python 语言开发，并通过 Pycharm 平台编译运行。如图 5 所示，以浙江省台州市市府大道与东环大道交叉路口实测交通数据为测试样本，构造了道路指示灯仿真环境。运用基于 Deep Q Networks 的道路指示灯控制方法进行仿真计算。基于 Deep Q Networks 的道路指示灯控制方法中的参数预设如表 1 所示。

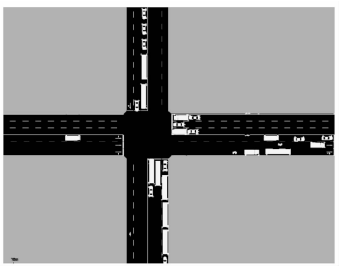

图 5 道路指示灯仿真场景

表 1 方法参数预设

参数	赋值
训练步长 K	7 000
记忆库 M	3 600
目标网络更新率 α	0.006
折扣因子 γ	0.84

针对道路指示灯控制问题，运用基于 Deep Q Networks 的道路指示灯控制方法进行仿真计算，方法效果如图 6 所示。由结果可知，随着训练次数的增加，训练误差不断减少，价值函数趋于稳定，方法收敛效果较好、稳定性较强，能够适用于道路指示灯自主控制问题。

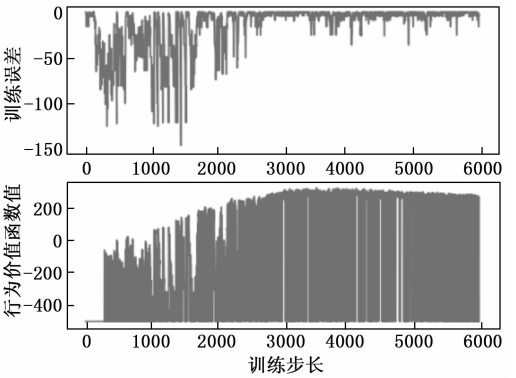

图 6 方法实验效果

如图 7 所示，为运用本文方法与固定时间控制模式、Webster 控制模式的仿真实验结果对比图。通过仿真结果对比可知，本文方法能够获得更好的道路指示灯控制策略，在交通路口车辆队列长度、等待时间等方面优化效果更为明显，有效地减少车辆停留时间、车辆延误，从而有效缓解高峰时期的交通拥堵现象。

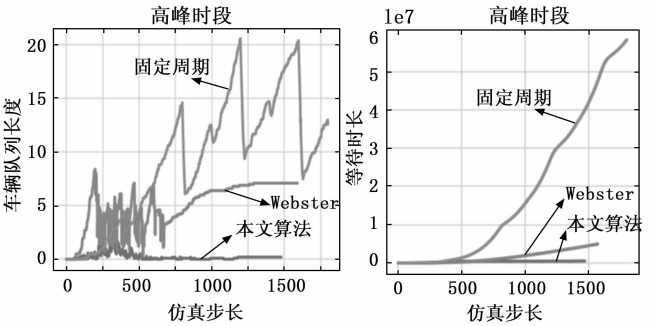

图 7 方法对比结果

5 结束语

针对当前城市道路拥堵、十字路口车辆通行效率低的问题，本文提出了基于 Deep Q Networks 的道路指示灯控制方法。首先，对道路指示灯控制问题进行描述，分析常用方法，聚焦十字路口道路指示控制的关键。然后，以状态、行动和奖励三要素为主构建了道路指示灯控制的强化学习模型，满足道路指示灯控制特征，提升了 Deep Q Networks 的适用性。最后，提出了基于 Deep Q Networks 的道路指示控制模型和方法流程，实现了道路指示灯的自主控制。基于城市道路数据的仿真计算，并与固定控制模式、韦伯斯特控制模式仿真比对，验证了论文方法对道路指示控制问题的适用性和优越性，为交通信号智能化控制提供了新的思路与途径。

参考文献：

[1] 白　轩，李　晋，张小虎，等. 基于线性回归预测的城市轨道交通车地无线通信性能提升方法研究 [J]. 计算机测量与控制，2020，28 (10)：145 - 150.

[2] 廖志斌，龙广钱，陈启新. 城市轨道交通车地无线通信系统性能评估平台的研究与实现 [J]. 计算机测量与控制，2020，28 (10)：247 - 250.

[3] 万　琴，朱晓林，肖岳平，等. 面向复杂城市交通场景的一种实时车道线检测方法 [J]. 计算机测量与控制，2019，27 (9)：61 - 65.

[4] 王丽君，史二娜. 基于卷积神经网络的智能交通信号控制研究 [J]. 信息技术，2020，44 (10)：56 - 60.

[5] 刘佳佳，左兴权. 交叉口交通信号灯的模糊控制及优化研究 [J]. 系统仿真学报，2020，32 (12)：2401 - 2408.

[6] 王　鹏，李艳雯，杨　迪，等. 基于层级控制的宏观基本图交通信号控制模型 [J]. 计算机应用，2021，41 (2)：571 - 576.

[7] 张　轮，张希雨，夏　凡，等. 基于监督机制的城市交通信号多智能强化学习控制方法 [J]. 交通与运输，2020，36 (4)：86 - 91.

[8] Hamilton A，Waterson B，Cherrett T，et al. The evolution of urban traffic control：changing policy and technology [J]. Transportation Planning and Technology，2013，36 (1)：24 - 43.

[9] 李靖丰. 基于 WPD－PSO－ESN 的城市交通感应信号控制系统设计 [J]. 计算机测量与控制，2020，28 (8)：144 - 148.

[10] 王亚飞. THMR－V 平台上的智能交通监控系统设计与实现 [J]. 计算机测量与控制，2017，25 (7)：106 - 109.

[11] 张　磊. 基于北斗卫星的狭窄路段交通拥堵智能控制系统设计 [J]. 计算机测量与控制，2020，28 (4)：121 - 125.

[12] 童　林，官　铮，杨文韬，等. 区分交通流模式的混合服务路口信号控制策略研究 [J]. 控制与决策，2021，36 (6)：1509 - 1515.

[13] Hartenstein H，Laberteaux L. A tutorial survey on vehicular ad hoc networks [J]. IEEE Communications magazine，2008，6 (46)：110 - 109.

[14] E van der Pol. Deep reinforcement learning for coordination in traffic light control [D]. Netherlands：University of Amsterdam，2016.

(上接第 92 页)

常状态下连续运行 168 h，并在期间不定期进行全功能操作，观察系统运行情况，监测各总线上的丢帧率，每条 CAN 总线上的丢帧率应为 0。

2) 开展被测设备运行系统整体的试启动运行测试实验。各子系统按正式工况运行，被测设备无负载运行，测试监控系统各功能指标。

3.2　实验结果

经过两轮测试，监控系统各功能指标和技术指标正常稳定，报警延时 1 s，数据刷新周期 1 s，数据存储间隔 3 s，数据库每 24 小时产生 700 M 数据，各阀门控制及连锁功能正常稳定，CAN 总线丢帧率为 0，现场站与上层客户端数据刷新无延时，满足启动运行条件。

4　结束语

文章通过对某设备运行试验监控系统的架构设计、专用检测装置研制、数据接口设计、数据存储与查询设计等内容的阐述，介绍了基于 Rockwell 自动化产品构建大规模考核试验监控系统的关键技术，该监控系统除了实现显示、控制、分析、报警、存储、查询等功能外，实时性、稳定性好，人机交互友好程度高，为新型设备的研制及试验提供了可靠保障。同时也为其它大规模运行监测系统的设计提供了借鉴。

参考文献：

[1] 陈志平. 现场总线技术及工业控制网络技术 [M]. 北京：电子工业出版社，2008.

[2] 孙传友，孙晓斌，汉泽西，等. 测控系统原理与设计 [M]. 北京：北京航空航天大学出版社，2002.

[3] 王亚刚，等. 基于 ROCKWELL 3 层网络架构的流水线控制系统设计 [J]. 控制工程，2012，19 (6)：935 - 938，943.

[4] 王　岩，等. 生产现场远程监控方法与系统 [J]. 现代制造工程，2020，1：113 - 117.

[5] 仇晓静. 基于 Web 的远程监控系统实施信息关键技术研究 [D]. 南京：南京理工大学，2013.

[6] 杜俊俐，王东云. 基于 C/S 与 B/S 混合架构的远程监控系统研究 [J]. 中原工学院学报，2003，14 (1)：1 - 3.

[7] 施圣杰，等. 基于 CAN 总线和 OPC 技术的多轴运动控制系统 [J]. 机械与电子，2015 (11)：54 - 58，61.

[8] 于　玲，杜向军，基于多总线 OPC 技术的组态监控系统设计 [J]. 天津职业院校联合学报，2014 (2)：70 - 73.

[9] 石先城，冯郁成，等. 基于 CAN 总线和 OPC 技术的分布式横向控制系统 [J]. 中国造纸，2015，34 (8)：44 - 48.

[10] 梁　庚，李　文. 基于分布式 OPC、组件连接件和 Web Server 的电站监控系统设计 [J]. 电力自动化设备，2011，31 (10)：134 - 138.

[11] 冯国川，等. 基于 VB 的新型报表系统研究 [J]. 自动化技术与应用，2010，29 (12)：31 - 34.