文章编号：2095-2163(2019)06-0164-04 中图分类号：U491.54 文献标志码：A

基于异步深度强化学习的城市智能交通控制方法

徐恩灶，朱海龙，刘靖宇，石晔琼，尹启天
（哈尔滨师范大学 计算机科学与信息工程学院，哈尔滨 150025）

摘 要：本文针对城市智能交通信号控制领域存在的控制效果差，算法收敛速度慢等问题，提出了一种基于异步优势行动者—评论者算法的深度强化学习的城市智能交通控制算法。首先抽象出交通路口的特征，输入到由神经网络构成的智能体，通过多个智能体异步学习，解决了传统方法控制效果不理想、训练耗时过长的问题。通过在开源交通模拟软件 sumo 上进行仿真，对比固定时间和传统方法控制的交通路口信号灯，不同的交通流量情况下的交通路口通行效率都有所提高。实验证明本文提出的方法可以有效解决城市交通路口信号灯控制问题。

关键词：智能交通；深度学习；异步强化学习

Urban intelligent traffic control method based on asynchronous deep reinforcement learning

XU Enzhu , ZHU Hailong , LIU Jingyu, SHI Yeqiong ,YIN Qitian
(School of Computer Science and Information Engineering, Harbin Normal University Harbin, China)

【Abstract】 The control effect in the field of urban intelligent traffic signal control is not satisfied enough, and the algorithm converges slowly. This paper proposes an urban intelligent traffic control algorithm based on the deep superior entrant-reviewer algorithm for deep reinforcement learning. Firstly, the characteristics of the traffic intersection are abstracted and input into the intelligent body composed of the neural network. Through the asynchronous learning of multiple intelligent bodies, the problem that the traditional method control effect is not ideal and the training takes too long is solved. By simulating on the open source traffic simulation software sumo, comparing the traffic time signal lights controlled by fixed time and traditional methods, the traffic efficiency of traffic intersections under different traffic flow conditions has been improved. The experiment proves that the method proposed in this paper can effectively solve the problem of traffic light control at urban traffic intersections.

【Key words】 intelligent transportation; deep learning; asynchronous reinforcement learning

0 引 言

随着经济的发展及科技的进步，城市中的机动车越来越多，交通拥堵问题逐渐开始显现。导致城市交通拥堵的一个重要原因是城市道路交叉口的交通信号灯调度不合理，因此需要合理的交通信号灯调度来地缓解城市交通路口的拥堵。城市交通控制系统目的在于更好地控制城市交通路口交通信号灯，利用现有交通网络道路的基础设施，在没有增加大量成本的情况下有效缓解城市交通路口的交通拥堵。但是，设计适合的城市交通控制系统仍然是当前智能交通领域的热点和难点。

本文提出的方法主要优点在于：

(1)使用异步多线程技术，有效利用计算机资源。

(2)所有实验都在国际主流的开源交通模拟软件 sumo 上进行了仿真实验，方法的可信性和可靠性得到了充分验证。

(3)通过多次实验选择了合适的超参数，使得控制的稳定性得到了提高，程序执行的时间减少。

1 研究现状和存在的问题

城市交通控制系统的研究大致经历了以下几个阶段，早期的城市交通控制系统主要建立在一些简化的交通流模型上，并假设短期内交通路口的流量不变[1]。这种人工设置固定时间的时序方法存在一些不足：这种方法严重依赖于交通路口调度人员的经验；固定时序的交通信号灯在面对交通突发状况时无法做出有效地应对。随着人工智能理论和智能控制技术的快速发展，出现了基于强化学习的城市交通控制方法。已成功用于除交通控制以外的许多应用[2]。对于交通控制问题，基于强化学习的方

基金项目：黑龙江省教育厅科学技术研究项目(12541240)。

作者简介：徐恩灶(1995-)，男，硕士研究生，主要研究方向：人工智能、智能交通。

收稿日期：2019-09-10

法通常将交通路口周围的交通流状态视为可观察状态，将信号时序计划的变化视为动作，并将交通控制的效果视为反馈。在经过特征提取之后，交通控制问题可以被视为传统强化学习问题并通过使用一些传统强化学习算法来解决。基于基本强化学习的方法考虑了孤立交通路口的信号时序。其中大多数都是使用 Q-Learning[3] 和 SARSA[4] 等经典算法，用于控制单个交通路口的交通信号灯。但是传统的基于强化学习的方法使用表格来记录和描述状态和动作之间的关系。因此，很难将其用于具有多个交通路口的城市交通控制问题，因为状态—动作空间的维度太大而无法学习。

深度学习作为人工智能研究的最新和成功的突破之一，已被引入并与强化学习方法相结合。深度强化学习的好处在于其能够通过使用比表格更有效的数据结构（深度神经网络）来快速学习和捕获状态和动作之间的关系。深度学习和强化学习的整合，就是广为人知的深度强化学习，已经成功解决如视频游戏[5]、围棋游戏[6] 以及许多其它问题。Li 等人最早提出了使用深度强化学习方法解决交通控制问题。在文章中，研究人员将这种方法应用于不同的情景，通过新的流量状态编码方法或使用不同的模型（如深度确定性策略梯度），这些方法也得到了改进。但是，现有的基于深度强化学习的城市交通控制方法在具有多个交通路口的场景中并不具有很好的控制效果。第一，一些深度神经网络（例如深度 Q 网络）用于模型状态和动作之间的关系不适合包含多个交通路口的交通控制问题。第二，当交通路口之间的相关性变高时，一些简单的奖励函数就无法很好地描述交通系统的状态。第三，一些用于训练基于深度强化学习的城市交通控制模型的算法，无法在解决方案空间探索和最优解决方案之间保持适当的平衡，这些算法收敛太慢而无法成为大规模城市交通信号控制问题的成功方案。

2　强化学习算法

2.1　异步多智能体强化学习算法

为了解决上述问题，本文提出一种利用异步优势行动者—评论者深度强化学习算法的城市智能交通控制系统。使用新的强化学习的奖励函数，对城市交通路口的信号灯进行自适应控制。本文提出的城市交通控制方法，不仅解决了多个交通路口之间的协作问题及强化学习状态空间的表达，并且有效控制方案的时间得到了降低，有效地提高了城市交通路网的路口通行效率。

2.2　异步优势行动者—评论者算法

本文提出的基于异步优势行动者—评论者（asynchronous advantage actor-critic，A3C）方法的深度强化学习，较好地解决了系统的深度强化学习的收敛速度慢、学习效果差等问题。在异步深度强化学习方法中，A3C 在各类动作空间的任务控制上表现最佳[7-10]。其合并了以值函数为基础（Q learning）和以动作概率为基础（Policy Gradients）两类强化学习算法[11]。A3C 具有基于奖励值的优化模式和对高维数据的快速优化决策能力。算法原理如图 1 所示。

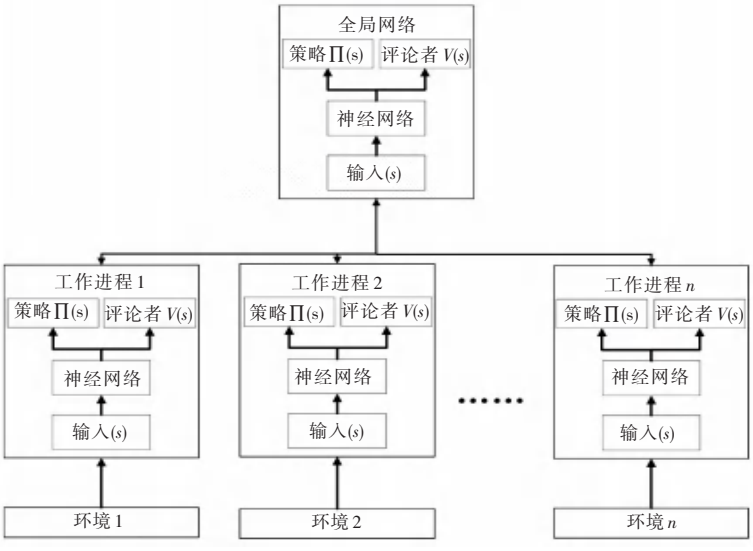

图 1　异步优势行动者-评论者算法原理

Fig. 1　Asynchronous advantage actor-commentator algorithm principle

2.3 场景介绍

交通路网的示意如图 2 所示。交通网络由 2 条南北方向的道路和 2 条东西方向的道路组成,每条道路长 500 m,这 4 条道路构成了 4 个交通路口。每条道路都是双向四车道。将交通路网的 4 个路口的等待车辆数量作为一个一维数组,输入神经网络。

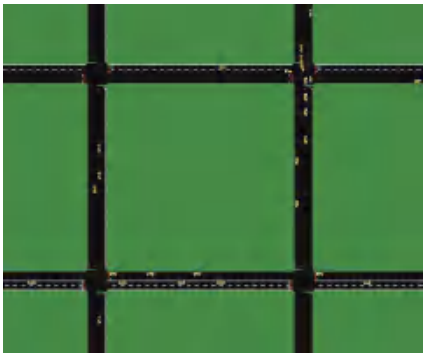

图 2 交通路网

Fig. 2 Traffic network

动作集合:每个交通路口的车辆有 4 种状态,南北直行、东西直行、南北左转、东西左转。右转总是被允许。场景如图 3、图 4 所示。

图 3 交通路口信号灯示意图

Fig. 3 Schematic diagram of traffic intersection lights

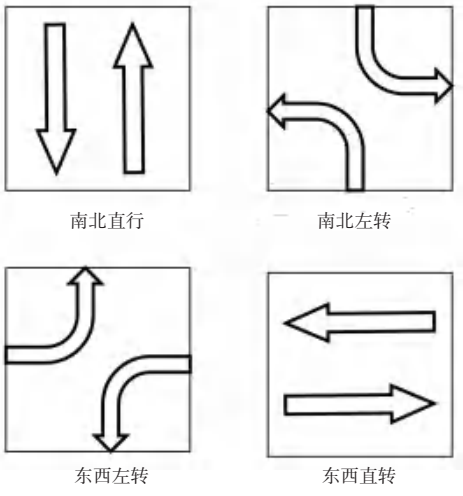

图 4 交通路口车辆转弯的方向

Fig. 4 Direction of turning in traffic intersections

2.4 算法设计

控制交通信号的智能体是由一个深度神经网络构成。这个神经网络是一个全连接神经网络,结构为 4 * 10 * 20 * 10 * 4,。输出层有 4 个神经元,每一个神经元对应着一个交通路口,每个神经元有 4 种状态,对应着交通路口的 4 种状态。

学习算法:使用异步优势行动者—评论者算法将异步方法引入到深度强化学习中,代替经验重放机制;利用多线程技术使多个模型同时训练打破数据间的相关性,提升算法的学习效果、学习速度和学习稳定性。

算法过程如下:

设公共部分的 A3C 神经网络结构,对应参数 θ, w,本线程的 A3C 神经网络结构,对应参数 θ', w',全局共享的迭代轮数 T,全局最大迭代次数 T_{max},线程内单次迭代时间序列最大长度 T_{local},状态特征维度 n,动作集 A,步长 α, β,熵系数 c,衰减因子 γ,探索率 ε。

输入:公共部分的 A3C 神经网络参数 θ, w

(1) 更新时间序列 $t = 1$;

(2)重置 Actor 和 Critic 的梯度更新量:$d\theta \leftarrow 0$, $dw \leftarrow 0$;

(3)从公共部分的 A3C 神经网络同步参数到本线程的神经网络:$\theta' = \theta$, $w' = w$;

(4) $t_{start} = t$, 初始化状态 s_t;

(5)基于策略 $\pi(a_t \mid s_t; \theta)$ 选择出动作 a_t;

(6)执行动作 a_t 得到奖励 r_t 和新状态 s_{t+1};

(7) $t \leftarrow t + 1$, $T \leftarrow T + 1$,

(8)如果 s_t 是终止状态,或 $t - t_{start} == t_{local}$,则进入步骤(9),否则回到步骤(5);

(9)计算最后一个时间序列位置 s_t 的 $Q(s,t)$:

$$Q(s,t) = \begin{cases} 0 & terminal\ state \\ V(s_t, w') & none\ terminal\ state, bootstrapping, \end{cases}$$

(10)for $i \in (t-1, t-2, \ldots, t_{start})$:

①计算每个时刻的 $Q(s,i)$:

$$Q(s,i) = r_i + \gamma Q(s, i+1),$$

②累计 Actor 的本地梯度更新:

$$d\theta \leftarrow d\theta + \tilde{\nabla}_{\theta'} \log \pi_{\theta'}(s_i, a_i)(Q(s,i) - V(S_i, w')) + c\,\tilde{\nabla}_{\theta'} H(\pi(s_i, \theta')),$$

③累计 Critic 的本地梯度更新:

$$dw \leftarrow dw + \frac{\partial (Q(s,i) - V(S_i, w')^2)}{\partial w'},$$

(11)更新全局神经网络的模型参数:

$$\theta = \theta - \alpha d\theta,\ w = w - \beta dw.$$

(12)如果 $T > T_{max}$，则算法结束，输出公共部分的 A3C 神经网络参数 θ，w，否则进入步骤(3)。

3　实验结果与总结

实验结果见表 1，使用异步优势行动者—评论者学习算法的城市智能交通信号灯，比传统的强化学习和固定时间的交通信号灯的控制效果有了明显的提升。而且本文提出的算法具有很好的适应能力，在不同的通行量的情况下，算法的执行效果都很好。表明算法具有很好的鲁棒性。

表 1　算法对比结果

Tab. 1　Comparison test results

通行数量/小时	对比算法	效率提升/%
3 000	q-learning	12.14
	固定时间	25.35
3 500	q-learning	11.96
	固定时间	24.15
4 000	q-learning	11.24
	固定时间	24.52

4　结束语

本文通过分析传统方法，提出了一种基于异步深度强化学习算法的城市智能交通控制方法，该方法在具有多个交通路口的城市交通路网控制方面不仅控制效果得到了提高，同时充分利用了计算机资源，使得算法在控制效果提升的同时算法收敛所用的时间相比于传统的方法也有了减少。但是该方法还存在一些可以改进的空间，比如在更为复杂的大规模城市交通路网、具有行人的交通场景等，是下一步的研究目标。

参考文献

[1] HUNT P B, ROBERTSON D I, BRETHERTON R D, et al. SCOOT-a traffic responsive method of coordinating signals[J]. Tech. Rep., 1981.

[2] ABDULHAI B, KATTAN L. Reinforcement learning: Introduction to theory and potential for transport applications,[J]. Canadian Journal of Civil Engineering, 2003, 6(30):981-991.

[3] EL-TANTAWY S, ABDULHAI B, ABDELGAWAD H. Design of reinforcement learning parameters for seamless application of adaptive traffic signal control [J]. Journal of Intelligent Transportation Systems, 2014, 18(3): 227-245: 2014.

[4] WATKINS C J, DAYAN P. Q-learning[J]. Machine learning, 1992, 8(3/4): 279-292.

[5] THORPE T L. Vehicle Traffic Light Control Using SARSA, Online]. Available: citeseer. ist. psu. edu/thorpe97vehicle. html, Tech. Rep., 1997.

[6] SHOUFENG L, XIMIN L, SHIQIANG D. Q-Learning for adaptive traffic signal control based on delay minimization strategy [C]// Networking, Sensing and Control, 2008. ICNSC 2008. IEEE International Conference on. IEEE, 2008, 687-691.

[7] MNIH V, BADIA A P, MIRZA M, et al. Asynchronous methods for deep reinforcement learning[C]//Proceedings of the International Conference on Machine Learning, New York, USA, 2016: 1928-1937.

[8] GENDERS W, RAZAVI S. Evaluating reinforcement learning state representations for adaptive traffic signal control [J]. Procedia Computer Science, 2018, 130: 26-33.

[9] HUSSEIN A, ELYAN E, GABER M M, et al. Deep imitation learning for 3D navigation tasks [J]. Neural Computing and Applications, 2018, 29(7): 389-404.

[10] PEROT E, JARITZ M, TOROMANOFF M, et al. End-to-End Driving in a Realistic Racing Game with Deep Reinforcement Learning [C]//Computer Vision and Pattern Recognition Workshops, Honolulu, USA, IEEE, 2017: 474-475.

[11] KAELBLING L P, LITTMAN M L, MOORE A W. Reinforcement learning: an introduction [J] . IEEE Transactions on Neural Networks, 2005, 16(1): 285-286.

(上接第 163 页)

标识别与跟踪等热门领域的需求，为进一步实现核环境下自主式导航机器人提供了基础。

参考文献

[1] 蔡自兴. 抗核辐射机器人的开发应用与警示[J]. 机器人技术与应用, 2011(3):24-27.

[2] 杜树标, 蒋韦韦, 丁泮. 核环境机器人现状及关键技术分析[J]. 兵器装备工程学报, 2016, 37(5):94-97,103.

[3] LOWE D G . Distinctive image features from scale-invariant keypoints[J]. IJCV, 2004, 60: 91-110.

[4] BAY H, ESS A, TUYTELAARS T, et al. SURF: Speeded up robust features (SURF) [J]. Computer Vision and Image Understanding, 2008,110(3):346-359.

[5] 索春宝, 杨东清, 刘云鹏. 多种角度比较 SIFT、SURF、BRISK、ORB、FREAK 算法[J]. 北京测绘, 2014(4):23-26,22.

[6] 王金龙, 周志峰. 基于 SIFT 图像特征提取与 FLANN 匹配算法的研究[J]. 计算机测量与控制, 2018,26(12):175-178.

[7] HAJEBI K, ABBASI-YADKORI Y, SHAHBAZI H, et al. Fast approximate nearest-neighbor search with k-nearest neighbor graph[C]// Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence. Catalonia, Spain: AAAI, 2011:1312-1317.

[8] FISCHLER M A, BOLLES R C . Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981, 24(6):381-395.

[9] 王卫兵, 白小玲, 徐倩. SURF 和 RANSAC 的特征图像匹配[J]. 哈尔滨理工大学学报, 2018, 23(1):117-121.

[10] 张继明,宋顾周,王群书,等. 辐射图像脉冲去噪方法[J]. 光子学报, 2010, 39(11):2107-2111.

[11] 陈敏, 汤晓安. SIFT 与 SURF 特征提取算法在图像匹配中的应用对比研究[J]. 现代电子技术, 2018, 41(7):41-44.