文章编号:1000-6788(2006)08-0136-05

基于 Q-学习的交通信号控制方法

承向军[1],常歆识[2],杨肇夏[1]
(1. 北京交通大学交通运输学院,北京 100044;2. 北京交通大学软件学院,北京 100044)

摘要: 为了减少车辆通过路口时的延误,采用 Q-学习方法对智能体控制的单路口进行信号配时的优化,在模糊控制规则集的基础上,通过 Q-学习来改进控制规则的组合,从而达到改善信号控制效果的目的. 仿真实验的结果表明,基于 Q-学习的信号控制方法优于定时控制、感应式控制和基于遗传算法的信号控制方法. 研究说明,基于 Q-学习的信号控制方法适合城市交通控制.

关键词: Q-学习;智能体;交通信号控制
中图分类号: U121　　**文献标识码**: A

A Traffic Signal Control Method Based on Q-Learning

CHENG Xiang-jun[1], CHANG Xin-shi[2], YANG Zhao-xia[1]
(1. Traffic and Transportation School, Beijing Jiaotong University, Beijing 100044, China; 2. Software School, Beijing Jiaotong University, Beijing 100044, China)

Abstract: In order to reduce the delay of vehicles passing intersection, we optimize the signal timing of agent controlled intersection by Q-Learning method. On the basis of fuzzy rule set for signal control, we improve the effect of signal control through optimizing the combination of control rules with Q-Learning. The result of simulation illustrates that the signal control method based on Q-Learning is better than fixed-time control, actuated control and signal control based on genetic algorithms. The result of this research indicates that the signal control method based on Q-Learning adapt to the urban traffic control.

Key words: Q-Learning; agent; traffic signal control

1 引言

路网的控制是一个非常复杂的问题,单个路口是交通网中最基本的结点,它的交通信号控制是路网控制的基础,解决好单路口的交通控制问题对解决整个路网的交通问题具有重大的意义. 随着社会经济的迅速发展,道路越来越拥挤,单纯使用基础建设手段来解决问题越来越困难,智能交通控制系统手段是一种较好的方法. 通过对单路口交通信号控制的研究,可以发现城市交通控制中的一些基本规律,有助于寻找对整个路网的交通信号进行有效控制的新方法.

陈洪[1]采用模糊控制方式对单路口进行交通控制,并通过模拟实验证明了这些方法的有效性,但它们不具有自学习能力. Ella Bingham[2]将交通路口模糊控制中的参数通过神经网络计算得到,改进了模糊控制的效果. Danko A. Roozemond[3]提出了以智能体为控制单元的交通控制模型,通过实际车流数据修正预测值来改善配时方案,但智能体不具有学习能力.

Baher Abdulhai[4]等、马寿峰[5]等使用具有 Q-学习能力的智能体控制单个路口,模拟实验表明这些方法优于定时控制. 两者都是利用了 Q-学习具有的强化学习特点,不必估计环境模型,不必知道控制行为是否正确,直接优化一个可迭代计算的 Q 函数,通过函数修正量来改进控制指标,达到学习的效果,只是两者选取的 Q 函数的构成项有所不同.

本文阐述了一种单路口交通信号控制方法,利用 Q-学习来实现智能体的自学习功能,给出了具体的

收稿日期:2005-07-12

作者简介:承向军(1968 -),男,北京交通大学交通运输学院,博士,讲师,研究领域:城市交通控制、智能交通、交通仿真.

Q-学习算法,并做出了仿真实现,在与其他几种方法进行对比后,证明在路口交通量较大时优化效果更明显.

2　基于 Q-学习的单路口交通信号控制

单路口的信号控制由一个控制智能体来完成,该智能体采用改进的反应模型,它由感知系统、信号控制规则集、信号控制行为系统等三个主要部分组成(见图 1).

信号控制智能体的感知系统通过检测装置获得车辆到达信息,并从内部获得当前信号显示状态信息. 信号控制规则集存储针对各种车辆到达情况的信号控制规则,这些规则是信号控制决策的依据. 信号控制行为系统包括信号控制智能体可能采取的所有行为,这些行为直接作用于信号,维持或改变当前的信号显示状态. 信号控制智能体根据感知系统获得的车辆到达数据和信号显示状态信息,从信号控制规则集中选取与之对应的控制规则,经过判断从信号控制行为系统中选出当前应采取的行动并执行.

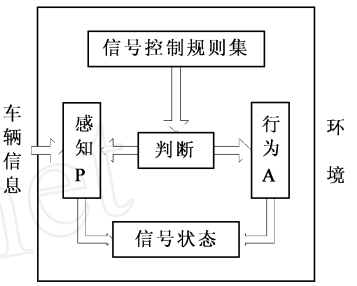

图 1　单路口信号控制智能体结构示意图

2.1　Q-学习基本原理

强化学习[6]是基于马尔可夫决策过程(Markov Decision Process)的,智能体可感知到其环境的不同状态集合 S,并且有它可执行的动作集合 A,在每个离散事件步 t,智能体感知到当前状态 S_t 选择动作 a_t 并执行,环境响应智能体并给出回报 $r_t = r(s_t, a_t)$. 强化学习中一个重要的里程碑就是 Q-学习算法,它是由 Watkins[7] 于 1989 年提出来的.

Q-学习是对状态动作对的值函数进行估计以求得策略. Q-学习中最简单的一种形式为单步 Q-学习,其 Q 值的修正公式如下:

$$Q(S_t, a_t) \leftarrow Q(S_t, a_t) + \alpha[r_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, a_t)]. \tag{1}$$

2.2　单路口交通信号控制规则集

根据交通控制经验和交通控制一般规律,制定交通信号控制规则集. 针对路口的 n 种不同的路口交通状态,应该产生与之一一对应的 n 条规则,即对于每个特定的路口交通状态都有一条控制规则,规定在该状态下应采取的信号控制行为.

信号控制行为通常包括三种,即:继续当前的相位转换,延长当前绿灯相位的时间 Δ 秒,结束当前的绿灯相位. 当分别用 a_0、a_1、a_2 表示这三种行为时,则有信号控制行为集 $A = \{a_0, a_1, a_2\}$.

信号控制过程就是每隔 Δ 秒的时间对路口交通状态进行一次判断,根据路口交通状态和相应的控制规则采取控制行为. 针对不同交通状态和信号显示状态采取的控制行为组成的信号控制规则集可描述为

$$R = \{r_i \mid p_i \Rightarrow a_l, i = 1, 2, \cdots, n; l = 0, 1, 2; p_i \in P, a_l \in A\}. \tag{2}$$

式中,r_i 表示第 i 条规则. a_l 表示下标为 l 的行为. $p_i \Rightarrow a_l$ 表示当路口的交通状态为 p_i 时,采取 a_l 行为.

控制信号变化的控制规则集分为不变规则集和可变规则集. 前者根据交通规律和经验,确定特定交通状态下采取固定的信号控制规则;后者则可以在实际控制过程中,根据控制效果的好坏以学习的方式改进规则,修改在特定车辆到达状态下是否延长当前相位绿灯时间的决策变量的值.

不变规则集为 R_f,可变规则集为 R_v,信号控制决策行为为 D. D 为 0 时表示结束当前相位的绿灯时间,D 为 1 时表示延长当前相位的绿灯时间 Δ 秒,D 为 - 1 时表示继续当前的相位转换.

如果不变规则集中的规则数为 c_f,可变规则集中的规则数为 c_v,则 $c_f + c_v = n$.

以 C_Δ 表示在 Δ 秒内到达的车辆数,则

$$C_\Delta = \sum_{k=1}^{m} C_{\Delta,k}. \tag{3}$$

式中 $C_{\Delta,k}$ 为在 Δ 秒内相位 k 所控制的车道在一定距离内到达的车辆数.

则有

$$[C_{\Delta,1}, C_{\Delta,2}, \cdots, C_{\Delta,m}] \wedge D([s_\Psi, s_1, s_2, \cdots, s_m]) \rightarrow [s'_\Psi, s'_1, s'_2, \cdots, s'_m]. \tag{4}$$

式(4)表示,在 Δ 秒内,系统输入各个相位新增的车辆到达为[$C_{\Delta,1}$, $C_{\Delta,2}$, …, $C_{\Delta,m}$],此时,信号控制系统的状态为(s_Ψ , s_1 , s_2 , …, s_m),采取信号控制行为 D,信号控制系统的状态转变为[s'_Ψ , s'_1 , s'_2 , …, s'_m].

为了统计车辆的停车延迟,设考察的时间由 Γ 个长度为 Δ 秒的时间段组成,在第 i 个时间段的 Δ 秒内一直处于静止状态的车辆数为 C_{s_i},由静止状态变为行驶状态的车辆数为 C_{m_i},由行驶状态变为静止状态的车辆数为 C_{n_i}. 则在 Δ 秒内的总停车延迟时间为

$$T_i = \Delta C_{s_i} + \frac{1}{2}\Delta C_{m_i} + \frac{1}{2}\Delta C_{n_i} = \Delta\left(C_{s_i} + \frac{1}{2}C_{m_i} + \frac{1}{2}C_{n_i}\right). \tag{5}$$

式中, C_{s_i} 、 C_{m_i} 、 C_{n_i} 均为自然数.

于是,在全部考察的时间内,总停车延迟为

$$T = \sum_{i=1}^{\Gamma} T_i , \tag{6}$$

式中 Γ 为正整数.

显然,信号控制的目标函数应该为

$$\min T = \sum_{i=1}^{\Gamma} T_i . \tag{7}$$

2.3 基于 Q-学习的交通信号控制

根据车辆检测器范围内车辆数目的变动情况,以模糊聚类的方式分为十个级别, Z_1 = 无, Z_2 = 极少, Z_3 = 很少, Z_4 = 少, Z_5 = 较少, Z_6 = 中等, Z_7 = 较多, Z_8 = 多, Z_9 = 很多, Z_{10} = 极多. 车辆到达状态模糊聚类隶属度函数的取值方法是根据经验产生的,主要考虑在检测区最大检测车辆数确定的前提下,以 3 辆车作为每个模糊聚类级别的分类间隔,可以满足算法的精确度要求.

控制规则集 $R = \{S_c, S_n, D\}$ 中各元素的取值范围分别为:

S_c :当前相位的车辆到达状态,分为无、极少、……、极多等 10 种.

S_n :下一相位的车辆到达状态,分为无、极少、……、极多等 10 种.

D :信号控制决策变量,取值为 1 时,表示延长当前绿灯相位的绿灯显示,延长时间一般取 1～3 秒,这里采用 1 秒;取值为 0 时,表示结束当前绿灯相位的绿灯显示.

Q-学习的状态集为可变规则集,共有 2^{21} 个状态,每一个状态可采取的动作为 21 种.

通过仿真实验对 Q-学习的部分参数进行了调整,调整后的 Q-学习控制算法如下:

对每个 s , a 初始化表项

$$Q(s, a) = 0; \tag{8}$$

观察当前状态 s.

一直重复做:

以概率 $P(a_i \mid s)$ 选择一个动作 a_i 并执行它;

选择动作的概率为

$$P(a_i \mid s) = \frac{k^{Q(s,a_i)}}{\sum_j k^{Q(s,a_j)}}, \tag{9}$$

取 $k = e^{-3.0\times(0.993)^h}$, h 为已学习时间,以小时为单位,经过多次实验调整,选择 0.993 为退火系数;

接受到立即回报 r,回报 r 选用信号周期内绿灯侧通过的车辆数与红灯侧增加的车辆延误时间(以分钟为单位)的比值;

观察新状态 s';

对 $Q_n(s, a)$ 按照下式更新表项

$$Q_n(s, a) \leftarrow (1 - a_n) Q_{n-1}(s, a) + a_n[r + \gamma \max_{a'} Q_{n-1}(s', a')]; \tag{10}$$

其中 $a_n = \dfrac{1}{1 + \mathrm{visit}_n(s, a)}$ (s 和 a 为第 n 次循环中更新的状态和动作, $\mathrm{visit}_n(s, a)$ 为此状态-动作对在 n 次

循环中被访问的总次数,取 $\gamma = 0.99$).

3　仿真试验结果

为了验证 Q-学习控制算法的有效性,对基于 Q-学习、遗传算法、定时控制、感应控制四种控制方法进行了仿真实验,以单路口作为仿真对象,路口结构如图 2 所示,仿真软件以文献[8]中的单路口-多路口仿真系统为基础,添加了 Q-学习控制方法.

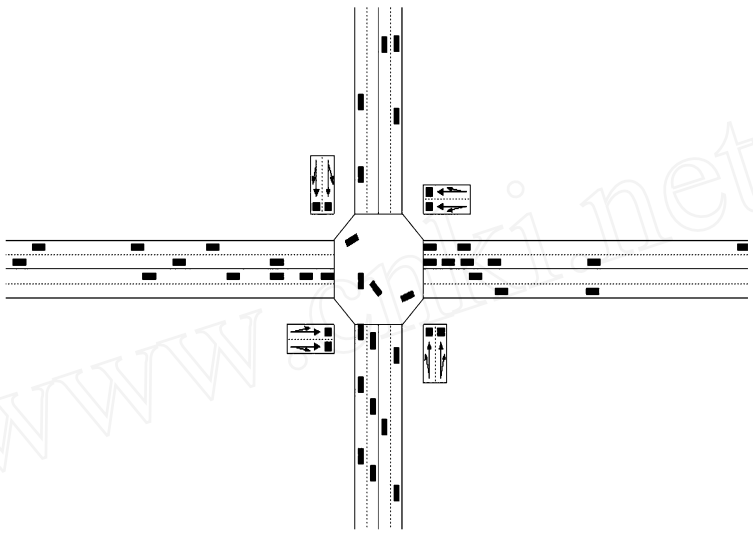

图 2　单路口仿真效果示意图

为了反映控制方法在不同环境条件下的效果,对单位时间内的车辆到达数目做出了多种规定,运行时间为 3 天(结果见表 1),并根据不同条件下对四种不同控制方法的指标做出了对比. 采用定时控制方式时,信号配时根据交通量以计算理论最优周期的韦伯斯特算法[9]得到;选用感应式控制方式时,最大绿灯时间为 35 秒,基础绿灯时间为 7 秒,单位延长绿灯时间为 3 秒;选择基于遗传算法的交通信号自学习控制方式时,最大绿灯时间为 35 秒,基础绿灯时间为 7 秒,单位延长绿灯时间为 1 秒,方案群体的规模为 16,淘汰方案数为 9,变异率为 0.015,迭代到第 81 代的控制效果;采用基于 Q-学习的控制方法时,最大绿灯时间为 35 秒,基础绿灯时间为 7 秒,单位延长绿灯时间为 1 秒,其他参数如 2.3 节所述,控制策略为运行 104 个信号周期之后趋于收敛时的信号控制效果.

表 1　几种控制方式的停车延迟时间

编号	交通量(辆/小时)	停车延迟时间(秒)			
		定时控制	感应控制	遗传算法	Q-学习
1	1080	19735.5	14213.2	13174.4	15212.6
2	1620	22174.8	16891.1	15707.0	16473.1
3	2160	26133.5	19086.7	17366.3	18066.0
4	2700	28716.2	24638.2	20629.6	19311.8
5	3240	30411.1	26752.3	22235.5	21011.5
6	3780	32475.1	30068.0	24338.7	22799.3
7	4500	35181.9	33974.9	26975.3	24718.1
8	5400	38777.9	38665.3	29250.9	26633.5

4　结论

从图 2 可以看出,通过对几种方式的仿真试验对比,我们发现其中感应控制优于定时控制,两种自学

习控制方法优于感应控制和定时控制. 随着交通量的增大,感应控制的效果逐渐与定时控制接近,当交通量趋于饱和时这两条曲线几乎相交. Q-学习控制方法在交通量较小时略差于感应控制,但随着交通量的增加,Q-学习控制方法的控制效果相对于感应式控制的改进程度更加明显.

从图 3、图 4 可以看出,Q-学习与定时控制和感应式控制相比改进明显,平均分别改进了 29.25 % 和 15.80 %,在交通量较大的时候,Q-学习明显优越于定时

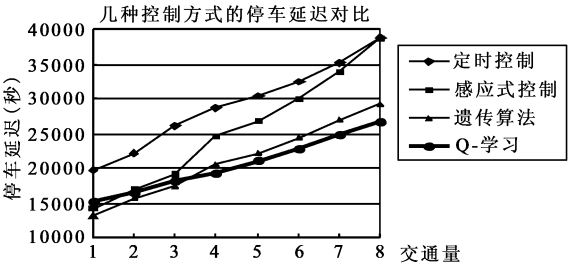

图 3 几种控制方式的停车延迟对比折线图

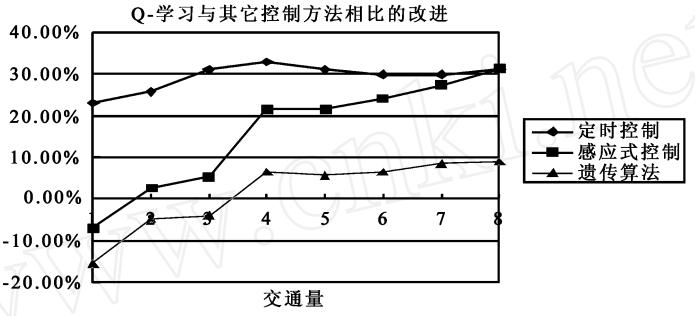

图 4 Q-学习与其它控制方法相比的改进程度

控制与感应式控制. Q-学习与遗传算法相比平均改进了 1.39 %,但在车流量最大的时候改进了 8.95 %. 实验结果证明在单路口中运用 Q-学习控制是一种较好的方法,优于传统控制方法.

下一步的工作是将单路口 Q-学习算法推广至多路口,并进一步改进算法结构.

参考文献:

[1] 陈洪,陈森发. 单路口交通实时模糊控制的一种方法[J]. 信息与控制,1997,26(3):227 - 233.
Chen Hong, Chen Senfa. A method for real-time traffic fuzzy control of a single intersection[J]. Information and Control, 1997, 26(3): 227 - 233.

[2] Ella Bingham. Reinforcement learning in neuro-fuzzy traffic signal control[J]. European Journal of Operational Research, 2001, 131: 232 - 241.

[3] Danko A. Roozemond. Using intelligent agent for pro-active, real-time urban intersection control [J]. European Journal of Operational Research, 2001, 131: 293 - 301.

[4] Baher Abdulhai, Rob Pringle. Machine learning based adaptive signal control using autonomous Q-learning agent [A]. Proceeding of the IASTED International Conference. Intelligent Systems and Control. Honolulu, Hawaii, USA. August 14 - 16, 2000: 320 - 327.

[5] 马寿峰,李英,刘豹. 一种基于 Agent 的单路口交通信号学习控制方法[J]. 系统工程学报,2002,17(6):526 - 530.
Ma Shoufeng, Li Ying, Liu Bao. Agent-based learning control method for urban traffic signal of single intersection[J]. Journal of Systems Engineering, 2002, 17(6): 526 - 530.

[6] Sutton R S. Introduction: The challenge of reinforcement learning[J]. Machine Learning, 1992, 8: 225 - 227.

[7] Watkins C J C H. Technical notes: Q-learning[J]. Machine Learning, 1992, 8: 55 - 68.

[8] 承向军,贺振欢,杨肇夏. 基于遗传算法的交通信号及其学习控制方法[J]. 系统工程理论与实践,2004,24(8):
Cheng Xiangjun, He Zhenhuan, Yang Zhaoxia. Machine-learning traffic signal control approach based on genetic algorithm[J]. Systems Engineering - Theory & Practice, 2004, 24(8):

[9] 翟润平,周彤梅. 道路交通控制原理及应用[M]. 北京:中国人民公安大学出版社,2002. 2.
Zhai Runping, Zhou Tongmei. The Theory and Application of Road Traffic Control [M]. The Publishing Company of Chinese People's Public Security University, 2002. 2.