

# 强化学习型匝道控制模型的研究

王兴举<sup>1,2</sup>, 宫城俊彦<sup>3</sup>

(1. 石家庄铁道大学 交通运输学院, 河北 石家庄 050043; 2. 河北省交通安全与控制重点实验室, 河北 石家庄 050043;  
3. 日本东北大学 信息科学研究科, 日本 仙台 9808578)

**摘要:**高速公路上的交通堵塞造成了道路利用效率低下,并伴随着能源消耗和环境污染问题,因此各种各样的高速公路控制方法应用于缓解交通堵塞。本文提出强化学习型匝道控制模型,该模型以交通流模拟为预测工具,以人工智能的强化学习为最优化选择模型,并具有一定的自主性、有记忆功能和性能反馈功能,且是一种动态的过程。应用 JAVA 针对不同的交通状态进行模拟再现,模拟结果表明匝道控制模型对于减少交通堵塞具有显著的效果。

**关键词:**匝道控制;强化学习;跟驰理论;车道变更;交通模拟

**中图分类号:**U491 **文献标识码:**A **文章编号:**2095-0373(2010)02-0104-05

## 0 引言

交通堵塞是动态的交通现象。局域化的车群堵塞是堵塞发生在公路某地的堵塞,表现为交通堵塞只在时间方向传播而同质交通堵塞表现为堵塞在时间和空间两方向传播。交通流的相变过程是一个十分复杂的过程,它是指交通流状态在不同交通相之间的转变过程<sup>[1]</sup>。外界施加的干扰、交通瓶颈是诱发交通相变的重要因素<sup>[2]</sup>。为了控制并减少交通堵塞,描述交通堵塞动态现象的交通流模拟模型是必要的。王、宫城提出了基于期望速度的一种车辆换道模型,模型根据驾驶员换道行为是否有换道的意愿和是否可能换道两个方面,提出了心理制约条件模型和物理制约条件模型,并 Treiber 和 Helbing 提出的智能驾驶员模型结合,提出了新的微观交通流模拟模型,并应用此模型再现同步态(Homogeneous congested traffic)和固定的局域化车群堵塞(Pinned localized clusters)等几种模式的高速公路合流部的交通堵塞现象<sup>[3-6]</sup>。本研究通过与交通模拟相结合提出一种智能型匝道控制模型,进而实现缓解高速公路合流部的交通堵塞的目的。

## 1 匝道控制模型

在高速公路管理中,有多种控制方法都可以改善交通流的状态,常用的控制方式有匝道控制、主线控制、通道控制,这些控制方式在实际中根据使用条件的不同,效果各异,既可以单独使用,也可以混合使用。根据国外经验,入口匝道控制是应用最广泛也最有效的一种缓解高速道路拥挤的交通控制形式。通过入口匝道控制,可以限制进入高速公路的车辆数量和流入高速公路的节奏,使高速公路主线交通流处于最佳状态<sup>[7-9]</sup>。

Papageorgiou 把匝道控制分为反应型匝道控制和预测型匝道控制<sup>[10]</sup>。反应型匝道控制可分为单点型反应控制和协调型反应控制。单点型反应控制利用匝道及其相邻路段的实时检测数据作为控制决策的基础。协调型反应控制是协调多个入口匝道的反应式控制方法。单点型反应控制有 DC(Demand-Capacity;需求-容量)、OCC(Occupancy;占有率)、ALNEA 等。DC 和 OCC 控制的原理是对匝道的上游或下游的占有率进行实时测量来估算匝道下游剩余容量,再来确定入口匝道的调节率。ALINEA 算法是一种基于

收稿日期:2010-04-14

作者简介:王兴举,男,1978 年出生,博士,讲师。从事城市交通解析,高速公路智能化、信息化管理和应急交通的研究。

基金项目:国家自然科学基金项目(51008201);教育部留学回国人员科研启动基金资助项目

闭环反馈控制的匝道控制策略,通过维持主线下游占有率在一个期望值附近为控制策略,实现主线通过交通量最大的目的<sup>[11-12]</sup>。从基于ALINEA的高速公路匝道控制策略在欧洲几个国家的应用情况看,该策略简单、高效且易于实施<sup>[13]</sup>。岩田·坪田·川嶋提出了一种基于交通流模拟的匝道控制模型,但缺少必要的目标控制函数<sup>[14]</sup>。

## 2 交通流微观模拟模型

### 2.1 车辆跟驰模型

跟驰理论研究的一个主要目的是试图通过观察各个车辆逐一跟驰的方式来了解单车道交通流的特性,并应用数学模型表达跟驰过程中发生的各种状态。1953年,pipes提出了如下描述车辆跟驰模型的基本等式<sup>[15]</sup>  $\ddot{x}_{n+1}(t) = a[\dot{x}_n(t) - \dot{x}_{n+1}(t)]$ ,  $\ddot{x}$ ,  $\dot{x}$ 以及 $x$ 定义了加速,车速和车辆相关点的距离。 $n$ 、 $a$ 是一个连续变量。在这个模型中,表现了跟驰车辆加速度只与被跟驰车辆和跟驰车辆速度差相关。模型假设了速度差导致的延误时间非常微小以至于可以被忽略。为了排除这个缺陷,Chandler提出了一个反应延误时间 $T$ (Chandler, 1958)的概念。因为跟驰车辆的加速度也受车速以及车距的影响<sup>[16]</sup>,Gazis, Herman 和 Rothery提出了以下更全面的车辆跟驰模型:根据实际测量结果,Newell提出改进的跟驰模型<sup>[17]</sup>。然而跟驰模型仍有以下缺陷:当前车不存在或离前车距离无限大时,无论车速和车速差距为多少,加速距离都为0;这意味着车辆没有跟随前车时车辆将保持初始车速。当车速没有差别时,无论车距为多少,加速距离为0。这表明了一种理想现象,即使离前车距离为0时,跟驰车辆也不会刹车,即使离前车距离很长,跟驰车辆也不会加速。为了解决上述问题,Treiber 和 Helbing 提出了智能驾驶者模型<sup>[18]</sup>。在这个模型中,导入了期望车速和最小安全距离。

### 2.2 车道变化模型

由于跟驰模型的单车道局限性不能表现合流部的车辆行为和交织区的合流行为,Gipps(1986)提出的车道变换决策结构模型、Koutsopoulos(1996)提出的离散模拟模型,Peter(2005)提出基于博弈论的换道模型等描述驾驶员的变换车道行为的车道变换模型能比较真实地反映了实际的交通现象<sup>[19]</sup>。王、宫城(2006)通过对合流部交通流的分析,根据高速公路和匝道车辆的速度、位置和心理因素,把合流行为分为自由合流和强制合流,并提出相应的数学模型描述合流换道现象,进而与车辆跟驰模型结合,提出描述合流部的交通流模拟模型。为建立适应一般道路的交通流模拟模型,提出基于期望速度的双车道的换道模型,并与 Treiber 和 Helbing 的智能驾驶者模型相结合,提出了基于期望速度的交通流模拟模型<sup>[3-6]</sup>。

## 3 强化学习型匝道控制模型

强化学习(Reinforcement Learning)是一种重要的机器学习方法。强化学习通过感知环境状态信息来学习动态系统的最优策略,通过试错法不断与环境交互来改善自己的行为,并具有对环境的先验知识要求低的优点,是一种可以应用到实时环境中的在线学习方式,因此在智能控制,机器学习等领域得到了广泛研究。强化学习中的一个重要里程碑就是Q学习算法,它是一种模型无关的强化学习算法。Sutton认为Q学习是瞬时差分算法的一种,称其为离策瞬时差分算法(off-policy TD),最优行动值的估计的更新依赖于各种“假设”的动作,而不是根据学习策略所选择的实际行动<sup>[20]</sup>。

为了解决匝道控制问题,提出基于TD强化学习的匝道控制模型

$$Q(qin_t, r_t) \leftarrow Q(qin_t, r_t) + a[qout + \gamma \max_r Q(qin_{t+1}, r) - Q(qin_t, r_t)] \quad (1)$$

如图1所示,  $qin$  为主线上游交通量;  $r$  为匝道流入交通量;  $qout$  为主线下游交通量;  $dm$  为主线上游密度;  $dr$  为匝道密度;  $vm$  为主线上游平均速度;  $vr$  为匝道平均速度。

令合流部驻留交通量  $q$  为

$$q = qin + r - qout \quad (2)$$

则合流部驻留密度  $\rho_L$  为

$$\rho_L = q_{t+1}/L \quad (3)$$

主线流入由  $qin_{t+1} \leftarrow qin_{t+1} + q$  进行更新。 $qin$  可由检测器获得,即为已知量;匝道流交通量  $r$  为控制选择策略; $qout$  为匝道控制策略  $r$  的报酬; $\rho_L$  为合流部的驻留密度; $L$  为合流部长度。

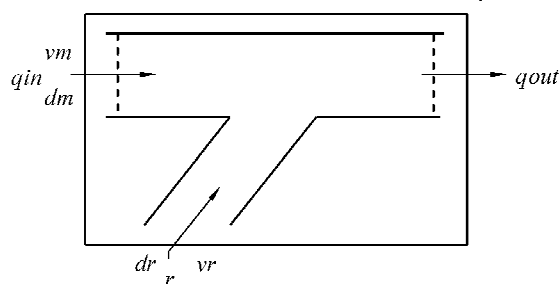


图1 强化学习型匝道控制手法概念图

图2 强化学习型匝道控制策略图

图2描述了强化学习型匝道控制方法的基本要素之间的关系。政策(policy):最优化条件为①主线不发生交通堵塞;②匝道不发生交通堵塞;③主线交通量最优化;根据王、宫城<sup>[4-5]</sup>交通模拟结果可知:当主线和匝道的密度大于15台/km以上,交通状态形成大的交通堵可能性很大。令主线和匝道的密度都小于15台/km为策略条件;报酬函数(reward function)为匝道流入交通量  $r$ ;价值函数(value function)为总流出交通量  $Q$ ;环境模型(model)为交通状态即本线流入交通量,驻留交通量,匝道流入交通量及相关速度密度指标如构成的整体。

主线上游交通量可由观测获得,作为状态变量;匝道流入交通量  $r$  为控制策略,作为选择变量; $qout$  为伴随着匝道控制策略的报酬。应用交通流模拟进行预测,并根据预测结果进行  $Q$  学习过程,进而进行控制决策选择。具体算法如图3所示。

该模型通过与匝道和交通状态的交互作用,根据过去获得的匝道控制经验信息,逐步改进控制系统自身的未来性能。具有一定的自主性:匝道控制系统的性能是自我改进的。一种动态过程:匝道控制系统的性能随时间而改变,性能的改进在与主线交通流量及主线交通状态反复作用的过程中进行。有记忆功能:匝道控制系统需要积累控制策略经验,用以改进其性能。有性能反馈功能:匝道控制系统需要明确它的当前主线道路交通量和密度与最适流量和密度之间的差距,施加改进控制策略<sup>[21-22]</sup>。

#### 4 强化学习型匝道控制模型的验证

仿真的目的是应用强化学习型匝道控制模型进行匝道控制,缓解合流部交通堵塞,从而验证强化学习型匝道控制模型的可行性及有效性。交通模拟中设定高速公路主线长1000m、匝道长200m、主线和匝道汇合区长100m。表1给出了实例A、B和C的主线道路交通量和有无强化学习型匝道控制的匝道流入交通量。

在实例B中,当主线交通量为1200台/h时,并不需要匝道控制;当主线交通量达到1600台/h时,匝道控制是必要的,匝道的控制策略是600台/h;当主线交通量持续增加到2200台/h时,最佳的控制策略是关闭匝道。图4是实例B中平均速度的时间分布图,如不进行匝道控制主线的平均速度降为4m/s,而进行匝道控制后的主线平均速度在25m/s上下波动。各个实例的控制效果如表2所示。在实例B中,以通过交通量为评价指标,交通量增加了280台/h,堵塞的缓和效果为15.22%,以平均行程时间为评价指标,较少了40s/台,堵塞的缓和效果为51.28%。

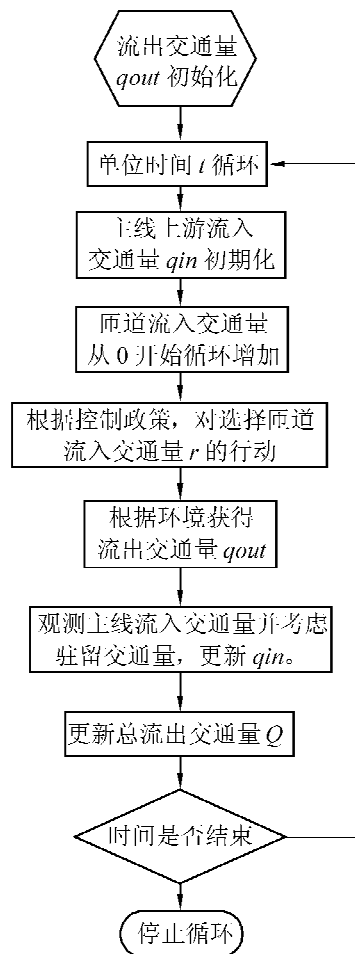


图3 匝道控制策略算法

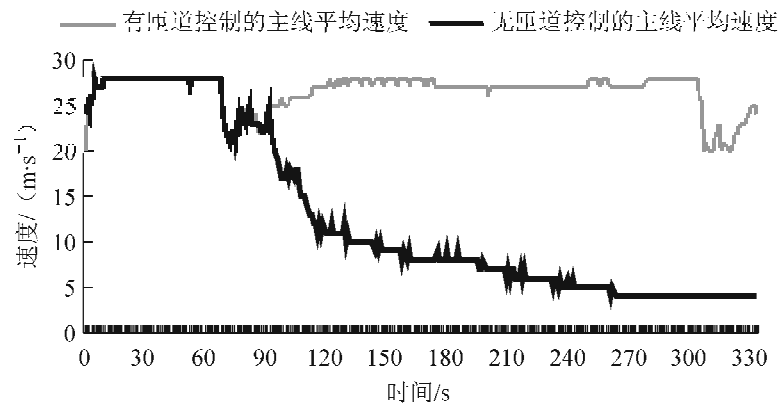


图 4 实例 B 中平均速度的时间分布

表 1 参数设定和匝道控制选择策略

| 实例 | 时间/min | 主线流入交通量/(辆·h <sup>-1</sup> ) | 匝道流入交通量/(辆·h <sup>-1</sup> ) |     |
|----|--------|------------------------------|------------------------------|-----|
|    |        |                              | 无控制                          | 有控制 |
| A  | 1.5    | 600                          | 800                          | 800 |
|    | 3      | 1 200                        | 800                          | 600 |
|    | 4.5    | 1 800                        | 800                          | 0   |
|    | 6      | 2 200                        | 800                          | 600 |
| B  | 1.5    | 1 200                        | 800                          | 800 |
|    | 3      | 1 800                        | 800                          | 600 |
|    | 4.5    | 2 200                        | 800                          | 0   |
|    | 6      | 1 600                        | 800                          | 600 |
| C  | 1.5    | 600                          | 800                          | 800 |
|    | 3      | 800                          | 800                          | 600 |
|    | 4.5    | 1 000                        | 800                          | 0   |
|    | 6      | 1 200                        | 800                          | 600 |

表 2 控制效果

| 实例                            | 无控制   | 有控制   | 变化量 | 堵塞缓解率/% |
|-------------------------------|-------|-------|-----|---------|
| A 交通量/(辆·h <sup>-1</sup> )    | 1 850 | 1 860 | 10  | 0.54    |
| A 平均行程时间/(s·辆 <sup>-1</sup> ) | 73    | 40    | -33 | -45.38  |
| B 交通量/(辆·h <sup>-1</sup> )    | 1 840 | 2 120 | 280 | 15.22   |
| B 平均行程时间/(s·辆 <sup>-1</sup> ) | 78    | 38    | -40 | -51.28  |
| C 交通量/(辆·h <sup>-1</sup> )    | 1 560 | 1 680 | 120 | 7.69    |
| C 平均行程时间/(s·辆 <sup>-1</sup> ) | 38    | 43    | 5   | 10.97   |

5 结论

匝道控制是缓和高速公路合流部交通堵塞的一种有效手段。本研究以能够表现时时刻刻变化的交通状态的交通模拟为工具,以强化学习为最优化选择模型,提出了动态智能型的强化学习型匝道控制模型。并应用交通模拟对模型的可行性进行验证,通过有无应用该模型的模拟结果进行对比,结果表明强化学习型匝道能够缓解高速公路合流部的交通堵塞。本文只验证了该模型的理论可行性,实际应用效果需要在今后的研究中进一步确认。

参 考 文 献

[1] Helbing D. High-fidelity macroscopic traffic equations[J]. Physical Review, 1995, 219: 391-407.  
[2] Kerner BS. The Physics of Traffic[M]. New York: Springer, 2004.  
[3] 王兴举,宫城俊彦. 希望速度を考慮したモデルによる渋滞波及現象のシミュレーション[C]//土木計画学研究・講演集. 東京:日本土木学会, 2005.  
[4] 王兴举,宫城俊彦. 渋滞波及現象モデルによるオンランプ区間での混雑現象分析[C]//交通工学研究発表会論文報

- 告書. 東京: 日本土木学会, 2006: 29-32.
- [5] Wang X J, Miyagi T, Ying J Q. A Simulation Model for Traffic Behavior at Merging Sections in Highways[ C]//Proceedings of Second International Conference on Innovative Computing, Information and Control, IEEE, 2007.
- [6] Wang X J, Miyagi T, Takagi A, et al. Analysis of the Effects of Acceleration Lane Length at Merging Sections By Using Micro-Simulations[ C]//Proceedings of The 7th International Conference of Eastern Asia Society for Transportation Conference, EAST, 2007.
- [7] 明神証, 坂本破魔雄, 岩本俊輔. 待ち行列を考慮したLP 制御[ C]//交通工学研究発表会論文報告書. 東京: 日本交通工程研究学会, 1975.
- [8] 松井寛, 佐藤佳朗. 都市高速道路の動的流入制御理論に関する研究[ C]//土木学会論文報告集. 東京: 日本土木学会, 1982(326): 103-114.
- [9] 秋山孝正, 佐佐木綱. ファジィ流入制御モデルを用いた交通制御方法の評価と検討[ J]. 土木学会論文集, 1991(413): 77-86.
- [10] Papageorgiou M, Kotsialos A. Freeway ramp metering: an overview[ C]//Proceedings of IEEE Conference on Intelligent Transportation System, IEEE, 2002: 271-281.
- [11] Papageorgiou M, H Haj-Salem, F Middleham. ALINEA Local ramp metering: summary of field results[ J]. Transportation Research Record, 1997, 1603: 90-98.
- [12] M Papageorgiou, H Haj-Salem, J-M. Blosseville. ALINEA: A Local feedback control law for on - ramp metering[ J]. Transportation Research Record, 1991, 1320: 58-64.
- [13] E Kenis, R Tebeos. Ramp metering Synthesis, Report of the Centrico Project[ R]. Brussels, Belgium: European Commission, 2001.
- [14] 岩田理史, 坪田幸政, 川嶋弘尚. 渋高速道路におけるランプ制御に関する比較研究[ C]//交通工学研究発表会論文報告書. 東京: 日本土木学会, 2006: 73-76.
- [15] L A Pipes. An Operational Analysis of Traffic of Traffic Dynamics[ J]. Journal of Applied Physics, 1953, 24(3): 274-287.
- [16] R E Chandler. Studies in Car Following[ J]. Oper. Res., 1958, 6: 165-184.
- [17] G F Newll. Nonlinear Effects In the Dynamic of Car Follow[ J]. Oper. Res. 1961, 9: 209-229.
- [18] M Treiber, A Hennecke, D Helbing. Congest traffic state in empirical observations and microscopic simulation[ J]. Review E., 2000, 62: 1805-1824.
- [19] Peter H. Modelling vehicle interactions in microscopic simulation of merging and weaving[ J]. Transportation Research Part C, 2005, 13: 63-74.
- [20] S S Richard, G B Andrew, 三上貞芳(訳), et al. 強化学習[ M]. 東京: 森北出版, 2000.
- [21] 王興拳, 宮城俊彦, 高木朗義. 強化学習型オンランプ制御手法に関する研究[ C]//土木計画学研究・講演集. 東京: 日本土木学会, 2007.
- [22] Wang X J, LIU B H, NIU X Q, et al. Reinforcement Learning Control for On Ramp Metering Based on Traffic Simulation[ C]//Proceedings of 2009 International Conference of Chinese Transportation Professionals. ASCE, 2009: 2071-2077.

## Reinforcement Learning Ramp Metering

Wang Xingju<sup>1,2</sup>, Miyagi Toshihiko<sup>3</sup>

(1. School of Traffic and Transportation, Shijiazhuang Tiedao University, Shijiazhuang 050043, China;

2. Traffic Safety and Control Laboratory of Hebei Province, Shijiazhuang 050043, China;

3. Graduate School of Information Sciences, Tohoku University, Sendai 9808578, Japan)

**Abstract:** Since the traffic congestion in a highway brings about an efficiency fall of road operation as well as an increase in energy consumption and environmental pollution, various kinds of traffic control have been considered for easing traffic congestion until now. In this paper, reinforcement learning model is introduced. By combining this model with a simulation model for describing the traffic flow behavior in the merging sections in highways, a novel reinforcement learning ramp metering model is proposed. By numerical simulation experiments, this model shows that the effect of the proposed control measure is effective in the highway.

**Key words:** ramp metering; reinforcement learning; car following model, lane change model, traffic flow simulation