Journal of Computer Applications
计算机应用,2014,34(2):601 - 604
ISSN 1001-9081
CODEN JYIIDU
2014-02-10
http://www.joca.cn

文章编号:1001-9081(2014)02-0601-04
doi:10.11772/j.issn.1001-9081.2014.02.0601

基于博弈学习的多 Agent 城市交通协调控制

郑延斌,王 宁*,段领玉
(河南师范大学 计算机与信息工程学院,河南 新乡 453007)
(*通信作者电子邮箱 wangning87@foxmail.com)

摘 要:交通路口中的各 Agent 之间的协调问题是一个博弈问题。在有限理性的基础上,利用博弈学习思想,构建多智能体(multi-Agent)博弈学习协调算法,利用此学习协调算法对出行者行为分析并修正,实现城市交通路口的畅通,进而达到区域、全局的交通优化。最后通过实例仿真验证其可行性。

关键词:有限理性;博弈学习;多智能体;协调算法
中图分类号:TP18 **文献标志码**:A

Multi-Agent urban traffic coordination control research based on game learning

ZHENG Yanbin, WANG Ning*, DUAN Lingyu
(*College of Computer and Information Technology, Henan Normal University, Xinxiang Henan 453007, China*)

Abstract: The coordination problem between Agents in traffic intersections is a gambling problem. On the basis of bounded rationality, this paper tentatively made use of game learning thought to build the multi-Agent coordinate game learning algorithm. This learning coordination algorithm analyzed travelers' unreasonable behavior and corrected it to realize the urban traffic intersections unimpeded, so as to achieve regional and global transportation optimization. At last, its feasibility is verified by means of an example and simulation.

Key words: bounded rationality; game learning; multi-Agent; coordination algorithm

0 引言

随着经济的高速发展,城市交通拥挤现象日益严重,如何有效地解决城市交通问题显得愈来愈突出,它直接影响城市经济的发展和居民生活的质量。解决交通拥堵问题的主要方法是通过建立交通流的数学模型,运用运筹学和控制理论来控制和优化整个交通系统,但由于交通系统的复杂性、交通流模型的局限性等因素,控制优化的效果并不理想。因此,随着智能控制的快速发展,一些先进的控制理论和方法,特别是智能控制的方法开始应用于城市交通控制[1-3]。城市交通中的每个出行者可以建模为一个 Agent,每个 Agent 的决策要受到其他 Agent 决策的影响,博弈论是描述这种相互影响的决策行为的最佳工具,因此将博弈论与多 Agent 结合应用于交通协调控制已经成为城市交通研究的热点[4-5]。周晶等[6]对公交网络系统的经营博弈问题进行了研究分析,以公交车车费作为经营者的决策变量,建立了经营者之间广义 Nash 均衡博弈模型,并将之转化成一个拟变分不等式问题再进行求解;郑长江等[7]对城市中无信号控制路段运用博弈论进行了分析;李静等[8]运用博弈论的概念与方法,剖析人们对公共道路和公交客源的利用,研究公交运营规模的相互影响,博弈结果证明了对公交运营进行宏观协调的必要性和重要性;黄园高等[9]对 N 个人合作博弈的 Nash 及演化均衡策略进行分析,通过实例运用生物动态复制理论验证演化均衡策略的有效性。Paissan 等[10]主要分析了在十字路口发生交互的出行者采用不同策略下的不同支付成本,通过动态模仿受教育水平高的出行者会采取合作行为,大大减缓了交通路口堵塞的可能,指出了对出行者进行教育的必要性。不论是智能控制还是结合博弈方法,都是在完全理性人假设的基础上,把所有的出行者都归结为一类进行研究(忽视了个体出行者之间的差异性),而事实上个体之间的差异是存在的,这也是在协调过程中必须考虑的问题。

为了处理个体之间的差异,首先介绍博弈学习理论的思想,在此基础上提出一种路口出行者之间的多 Agent 博弈学习协调方法,通过实例分析与仿真实验,证明该方法是可行的。

1 博弈学习

定理[11-12] 每个有限策略博弈,在连续的概率空间上,总存在混合纳什均衡。

博弈学习理论(theory of learning in games)是描述具有目标的有限理性参与者,如何通过逐步的调整策略来达到某一均衡点的。博弈学习是通过博弈学习主体的策略空间 S_t^* 来影射利益函数 U_{t+1}^* 的过程,通过学习(Learning),改变学习值 Learn(s,a),得到新的最佳 Nash 均衡值,即函数 $U_{t+1}^* = f(S_t^*)$,满足 $U_{t+1}^* \in BD_t^*$,$\forall t \geqslant t_0$,BD_t^* 表示参与主体学习后的最佳均衡点[13-14]。

假定博弈参与者可以选择两种行为方式合作(c)和对抗(d),r_i 表示选择行为 $i(i=c,d)$ 参与者在群体中的比例,

收稿日期:2013-08-12;**修回日期**:2013-10-14。
基金项目:河南省重点科技攻关项目(122102210086,132102210537,132102210538)。
作者简介:郑延斌(1964-),男,河南内乡人,教授,博士,主要研究方向:虚拟现实、多智能体系统、对策论; 王宁(1987-),男,河南邓州人,硕士研究生,主要研究方向:虚拟现实; 段领玉(1989-),女,河南新乡人,硕士研究生,主要研究方向:虚拟现实。

$N_i^0 \geq 1$ 群体中选择行为 i 的初始人数，$P_c(r_c^t, r_d^t)$ 表示在 $t+1$ 时刻选择行为 c 的概率，N_i^t 表示到 t 时刻选择行为 i 的人数，N^t 表示 t 时刻群体总的人数，则 $r_i^t = N_i^t/N^t$ 表示 t 时刻选择行为 i 的比例，$r_i^{t+1} = (N_i^{t+1})/(N^{t+1})$ 表示 $t+1$ 时刻选择行为 i 的比例，因此 $t+1$ 时刻选择行为 i 的群体的比例的期望为 $E(r_i^{t+1} \mid r_i^t) = r_i^t + (1/N^{t+1}) * (P_i^t - r_i^t)$，记向量 $r_c^t = (r_c^t, r_d^t)$，若 $r^* = (r_c^*, r_d^*)$ 是一个稳定的均衡点。则可以构造如下一个动态的博弈学习过程。

1）初始化，给定 N_c^0 和 N_d^0 的值以及利益函数 $U_c(r)$ 和 $U_d(r)$ 的表达式；

2）根据 t 时刻的情况，通过一定的数学方法计算 $P_c(r_c^t, r_d^t)$ 及 $P_d(r_c^t, r_d^t)$，转 3)；

3）根据 $E(r_i^{t+1} \mid r_i^t)$ 的表达式进行迭代，通过学习值 Learn(s,a) 的更新，预计可能出现的结果，根据利益函数 U_{t+1}^* 进行调整 r_i^t，转 4)；

4）判断，给定任意小正数 ε，若 $| r_c^t - r_c^* | < \varepsilon$ 且 $| r_d^t - r_d^* | < \varepsilon$，结束；否则转 2)。

通过上面的描述可以看出，博弈学习的主要特征如下：

1）与传统理论中完全理性人的假设不同，在博弈学习理论中，参与者是有限理性的或者自私的。受参与者的偏好、知识水平、所处的环境条件等各种因素的影响，参与者的认知、学习或推断能力是有限制的，即策略空间 S_t^* 是有条件限制的。

2）博弈学习理论针对的是一个动态变化过程。在每个时间段，参与者都根据自己所获得的信息，根据经验和自己的利益不断学习调整其策略和学习值 Learn(s,a)，即：$U_{t+1}^* = f(S_t^*)$。一般，参与者所收集的信息为：自己的行动历史、所有参与者的行动历史、其他参与者所采用的策略及收益，博弈学习理论是经过长时间的学习演化以后的一个长期结果。其中博弈学习 Learn(s,a) 的方法为（设群体规模为 S，选择合作的初始群体规模为 $S_{合作}^0$）：首先在一个交互的群体中辨别出每个 Agent 的状态，使这两个群体随机进行博弈学习选择行动策略，进而提出两个学习概率 P_x（选择合作的向选择竞争的学习的概率）和 P_{1-x}（选择竞争的向选择合作的学习的概率）。假定博弈群体采取合作方式会达到均衡，则经过博弈 P_{1-x} 会随着博弈学习越来越大，而 P_x 随着进化的开展越来越小 P_{1-x} 和 P_x 的初始值为 P_{1-x}^0 和 P_x^0，$S_{合作} = S * P_{1-x} + S_{合作}^0$，显然，结果选择合作的规模越来越大，最终就会达到较好的均衡。

3）在博弈学习理论中，针对不同类型的参与者可以有不同的目标，参与者对自己所获得的信息，根据自己的情况可以有不同的处理方式。这样可以使参与者在不同的博弈学习中将以不同的方式或方法对策略进行调整。

博弈学习提供了一种不同于传统理论的答案或解释。传统理论认为，Nash 均衡点是博弈规则，参与人是完全理性的，其收益函数是在共同知识库下产生的，由参与者通过分析得到结果。然而，在现实生活中对博弈参与者而言，完全理性的参与者是一个过高或者不切实际的要求[15]。例如，并非每一个参与者参与博弈过程时，都有足够的时间或能力进行充分的理性合理推断。针对此种缺陷，博弈学习理论就为 Nash 均衡点的产生或选择提供了一种与传统不同并且比较符合实际情况的解释。

2 基于博弈学习的 Multi-Agent 协调

定义 1 城市交通协调问题可以定义为一个博弈：$G = \{A, I, S, U\}$，其中 A 为博弈协调中决策主体，参与者的集合 $A = \{Agent_1, Agent_2, \cdots, Agent_n\}$，它是通过选择行动策略以最大化自己的效用水平；$I$ 是每个 Agent 拥有的信息，包括其他 Agent 的特征和行动策略的信息；S 为 Agent 的所有可能的策略或行动的集合，一个 Agent 的所有的可行策略称为它的策略空间，可以表示为 $S = \{S_1, S_2, \cdots, S_n\}$，每个参与者的策略可以形式化为 $S_i: Agent_i \rightarrow a_i (i = 1, 2, \cdots, N)$，其中 a_i 为参与者 $Agent_i$ 采取的行动，即 $a_i \in$ {东西直行，南北直行，东西左转，南北左转}；U 为利益函数，是指在既定策略组合条件下 Agent 的得失情况，即在一个特定的策略组合下参与者得到的效用水平。出行者 Agent 的利益函数是出行者通过路口延误时间最少，出行者 Agent 的利益函数是将时间、路段拥挤度、道路质量、耗油量作为衡量效用的综合指标。

定义 2 Nash 均衡的定义：对于定义 1 给出的博弈，假定策略组合为 $S^* = \{S_1^*, S_2^*, \cdots, S_n^*\}$，对任一博弈方 i 的策论 s_i^*，都是对其余博弈方策略的组合 $(s_1^*, \cdots, s_{i-1}^*, s_{i+1}^*, \cdots, s_n^*)$ 的最佳对策，也即 $u_i(s_1^*, \cdots, s_{i-1}^*, s_i^*, s_{i+1}^*, \cdots, s_n^*) \geq u_i(s_1^*, \cdots, s_{i-1}^*, s_{ij}^*, s_{i+1}^*, \cdots, s_n^*)$ 对任意 $s_{ij} \in S_i$ 都成立，则称 $(s_1^*, \cdots, s_n^*)$ 为 G 的一个纳什均衡。其中 S_i^* 为第 i 个 Agent 选择的策略，U_i 为第 i 个 Agent 的利益函数，S_i 为第 i 个 Agent 的策略空间。

2.1 多智能体协调结构

虽然每个 Agent 的决策是独立的，但是每一个 Agent 的决策会影响到其他 Agent 的决策，每一个 Agent 也受其他 Agent 决策的影响，因此一个 Agent 在做决策时，有必要考虑其他 Agent 可能采取的决策来决定自己的决策，各 Agent 之间必然会发生一定程度的冲突。为了进行冲突消解，必须进行相应的协调分析。为此设计了一个三层协调结构如图 1 所示，下层是出行者之间的协调，中间是路口之间的协调，上层是区域之间的协调。其目的是根据实际的交通情况，最大限度地使出行者都能尽快顺利通过各路口，它要求出行者之间进行协调，保证路口通畅，进而使得区域通畅，以达到全局的最优。

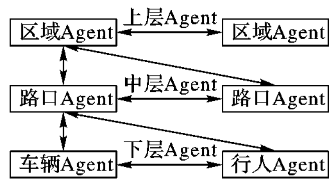

图 1 多 Agent 之间的 3 层协调结构

2.2 基于多 Agent 的协调过程

为方便起见，把路口的一个出行者 Agent 作为发起者，路口的其他出行者 Agent 作为参加者。出行者之间的协调可以概括为：发起者发出一个信号，每个参加者根据情况做出反馈，参加者再对这些反馈进一步处理，可能会出现多次提议—反馈—处理的过程。在协调开始前，每个出行 Agent 应该具有知识库，获取方式如图 2 所示。

基于多 Agent 的协调博弈过程描述如下：

1）假设车辆 $Agent_i$ 是发起者，行人 $Agent_j$ 是一个参加者，

发起者已经和参加者协商了 n 次，即 $F_i(n) \to F_j(n)$，其中 $F_i(\,)$ 表示 i 发起协商；$n \in (0, +\infty)$，表明车辆 $Agent_i$ 已发送给行人 $Agent_j$ n 个提议。

2）发起者 Agent 知道在此之前的每个 Agent 的动作。

3）车辆 Agent 使用博弈学习方法进而决定他下一步的动作。即 $\mathrm{Learn}(s,a) \to \mathrm{Learn}(s+1, a+1)$，$s$ 是每个 Agent 可能的状态终止或非终止，a 是行动策略集。

4）$F_i(n+1) \to F_j(n+1)$，行人 Agent 接收到来自车辆 Agent 的第 $n+1$ 次提议，就是车辆 Agent 的第 $n+1$ 次的动作。

5）车辆 Agent 更新博弈学习值 $\mathrm{Learn}(s,a)$，推断可能出现的情况，根据利益函数选择下一步的最佳动作。

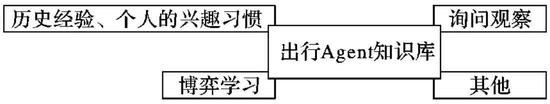

图 2　出行 Agent 知识库的获取方式

2.3　基于博弈学习的多 Agent 协调方法

假设在无信号灯管理的盲区，为了使出行者都可以顺利出行，把路口的多个出行者简化为参加者（车辆 Agent）和发起者（行人 Agent），具体博弈学习协调算法描述为：

初始化　对每个 s,a，初始化学习值 $\mathrm{Learn}(s,a)$，假设行人 a_j 获得车辆 a_i 的提议；

循环　直到 s_t 是终止状态，否则执行下面的操作。

1）行人 a_j 执行动作 a_t，调整学习值 $\mathrm{Learn}(s,a)$，得到新的学习值 $\mathrm{Learn}(s_{t+1}, a_{t+1})$ 并于车辆 a_i 发起博弈协调；

2）根据新的学习值 $\mathrm{Learn}(s_{t+1}, a_{t+1})$，调整策略空间 $S_t{}^*$，获得新的 Nash 均衡，得到新的利益函数 $U_{t+1}{}^* = f(S_t{}^*)$；

3）新的 Nash 均衡存在，车辆 Agent 和行人 Agent 向路口 Agent 反馈，进行下一个新的控制周期；否则，回到 1），寻找 Nash 均衡；

4）经过约定次数的博弈学习协调后仍找不到 Nash 均衡，协调失败，请求人工干预。

该算法的优点如下：

1）具有较好的寻找均衡点的特性，对初值不敏感，初值处置选择不当时，可以通过博弈学习后，仍能找到好的均衡点，满足控制的要求；

2）操作方便，不需要复杂的规则，只需通过一个简单的博弈学习，便可寻找到均衡点；

3）不仅适合理性群体优化，也适合有限理性群体优化。

2.4　实例分析

考虑有如图 3 所示的博弈，它们之间为有限理性重复博弈，其合作竞争的结构满足囚徒困境模型的要求。图中 R、S、T、P 表示不同的博弈策略对应的受益值。

		博弈方Ⅱ	
		合作(c)	叛变(d)
博弈方Ⅰ	合作(c)	R,R	S,T
	叛变(d)	T,S	P,P

图 3　合作—竞争博弈模型

博弈的四种结果：(c,c) 表示互相合作，(d,d) 表示互相叛变，(c,d) 或 (d,c) 表示博弈方之一单方面变节。假设该重复博弈有一大群有限理性的参与人，他们不可能一开始就找到最佳的策略（合作，合作），因此在参与人中有些是“合作”类型的，有些是“不合作”类型的，但这种类型不是事先给定的，而是根据参与人的习惯和得失在学习过程与策略调整中改变的，这就可以采用博弈学习协调算法。假设参与人群体中“合作”类型的比例是 x，则“不合作”类型是 $1-x$，群体的博弈是随机配对的，应用上面的分析结果，用来表示参与人的收益，于是“合作”类型参与人的收益为：

$$U_c = x * R + (1-x) * S$$

“不合作”类型参与人的收益为：

$$U_d = x * T + (1-x) * P$$

参与人的平均收益为：

$$U = x * U_c + (1-x) * U_d$$

于是通过求导可得“合作”类型的参与人比例的动态变化可表示为：

$$\frac{\mathrm{d}x}{\mathrm{d}t} = x * (1-x) * [x * (R-T) + (1-x) * (S-P)] \tag{1}$$

在图中的博弈中，每组策略组合下的收益都是给定的常量，因此 $\frac{\mathrm{d}x}{\mathrm{d}t}$ 仅为 x 的函数，于是式(1)可记为：

$$\frac{\mathrm{d}x}{\mathrm{d}t} = F(x) \tag{2}$$

要讨论该博弈的均衡策略，可先找到动态的稳定点，而后讨论其他影响稳定态的因素。

令 $F(x) = 0$，解得：

$x_1 = 0$；

$x_2 = 1$；

$x_3 = (S-T)/(R-S-T+P)$

于是得到 3 个稳定态，但是随着收益的不同取值，x_3 可能与 x_1 或 x_2 相等或者不存在，博弈就退化为只有两个稳定态。通过上面的分析，可以发现对于这样一组行人 — 机动车抢行模型：

表 1　行人—机动车抢行模型

行为	机动车抢行	机动车礼让
行人抢行	(6,6)	(4,10)
行人礼让	(10,4)	(10,10)

(6,6)，(10,10) 是均衡点，出行者为了追求利益最大化，尽管初始博弈时不一定都选择礼让，但是通过博弈学习，利用多智能体协作算法，最终大都逐渐会选择（礼让，礼让）组合。

3　仿真实验

在 Matlab2010 环境下基于以上的模型、实例和算法，对无交通信号的无人盲区十字路口的交通协调进行仿真，假设车辆到达服从泊松分布，路口车辆的限制阈值为 30，最大限制阈值为 40，以通过路口的顺利程度作为性能指标，1 表示最快通过，0 表示堵塞，越接近 1 表示性能越好，反之则性能较差。仿真通过对四种方法对比研究：

1）路口行人大都为了快速通过，盲目采用抢行通过的方式，抢行比例较高。

2）路口行人通过智能控制协调，选择通过方式。

3）路口行人通过简单博弈，通过判断选择出行方式。

4）路口行人通过博弈学习协调，选择合理的通过方式。

从图4仿真结果可以看出，在路口流量较少时，几种方法性能接近，在路口流量达到限制阈值时，普通方法会造成交通堵塞，而简单博弈方法和智能控制方法在车辆数目接近阈值时会变化速度较快，这两种方法的性能比较接近，博弈方法性能稍高。

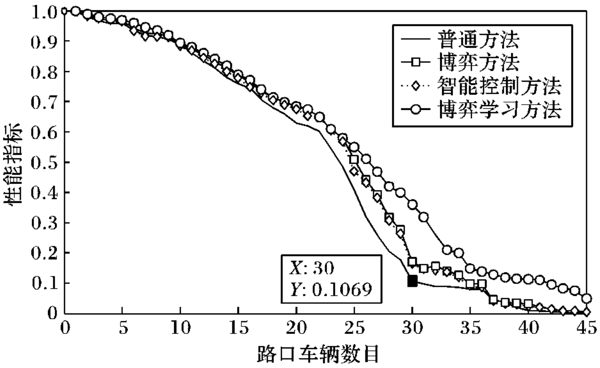

图4 采用不同的方法性能指标对比

从图4的对照可以看出，博弈学习协调方法性能比较好，它的性能在0.5左右，可以很好地保证交通的畅通。在车辆数目超过阈值接近限制流量时，前几种方法都容易出现交通拥堵现象，而博弈学习方法还有一定的协调能力，但在接近限制流量性能会快速下降，而且在车辆数目超过最大流量时博弈学习协调算法的作用就不明显了，就需要寻求路口协调、区域协调或其他道路管理控制手段，有时还需要人工干预。从图5中可以看出，利用博弈学习的方法，出行者会选择一种比较合理的出行方式，使得等待时间相对较短，达到节约时间的目的，最终实现路口的快速流通，缓解交通压力。另外，在接近阈值时平均等待时间比博弈方法等待时间短，并且趋于稳定，可以看出博弈学习算法的有效性。

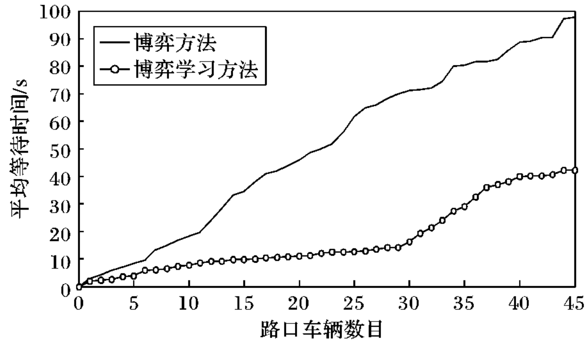

图5 博弈学习协调方法与博弈方法的对比

4 结语

城市交通系统的复杂性、随机性、分布式和巨大性等特点，多智能体博弈学习方法为其提供新的解决方法，在实现方面需要交通部门和政府加大宣传，通过博弈学习协调算法制定一定的奖惩措施，使出行者逐步都选择最优组合，提高出行效率。本文提出了出行者之间的多智能体博弈学习协调算法，并通过仿真实验对其有效性进行验证。从实验结果可以看出，采用博弈学习协调算法，在出行高峰期，可以使车辆更快通过路口，提高了路口的通行效率。而在现实中，如果遇到意外情况，如何进行协调控制就需要深入研究。

参考文献：

[1] ZHONG F. Traffic mode choice based on evolutionary game[J]. Technology & Economy in Areas of Communications, 2013, 15(1): 66－72.（钟芳. 公共交通出行方式选择的博弈分析[J]. 交通科技与经济, 2013, 15(1): 66－72.）

[2] HENRY Y K L, VICKY W K W, IVAN S K L. Immunity-based autonomous guided vehicles control[J]. Applied Soft Computing, 2007, 7(1): 41－57.

[3] BOROS E, GURVICH V. Perfect graphs, kernels, and cores of cooperative games[J]. Discrete Mathematics, 2006, 306(19/20): 2336－2354.

[4] TANG Q. Study on pedestrian and vehicle interference at signalized intersection based on cooperative game theory[D]. Beijing: Beijing Jiaotong University, 2010.（唐勍勍. 基于合作博弈的平面信号交叉口行人和机动车干扰研究[D]. 北京: 北京交通大学, 2010.）

[5] DONG T, TAN J. Analysis on game theory in traffic congestion[J]. Journal of Chongqing Normal University: Natural Science, 2012, 29(2): 99－102.（董甜甜, 谭建春. 博弈原理在解决交通拥挤中的分析[J]. 重庆师范大学学报: 自然科学版, 2012, 29(2): 99－102.）

[6] ZHOU J, XU Y. Generalized Nash managing game mode for transit network[J]. Journal of Systems Engineering, 2001, 16(4): 261－266.（周晶, 徐晏. 公共交通网络系统的广义 Nash 经营博弈模型[J]. 系统工程学报, 2001, 16(4): 261－266.）

[7] ZHOU X, ZHENG C. An analysis of road-grabbing at non-intersection crosswalks without signal control based on game theory[J]. Journal of East China Jiaotong University, 2012, 29(6): 65－69.（周雪峰, 郑长江. 基于博弈论的无控制路段人行横道处人车抢行分析[J]. 华东交通大学学报, 2012, 29(6): 65－69.）

[8] LI J, FAN B. Public transport operation scale analysis with game theory[J]. Urban Public Transport, 2003(2): 9－10.（李静, 范炳全. 公交运营规模的博弈分析[J]. 城市公共交通, 2003(2): 9－10.）

[9] HUANG Y, ZHOU J. Fare competition between highway and public transport[J]. Journal of Southeast University: Natural Science Edition, 2004, 34(2): 268－273.（黄园高, 周晶. 收费公路和公共交通之间的定价博弈分析[J]. 东南大学学报: 自然科学版, 2004, 34(2): 268－273.）

[10] PAISSAN G, ABRAMSON G. Imitation dynamics in a game of traffic[J]. The European Physical Journal B, 2013, 86(4): 1－6.

[11] GLICKSBERG I L. A further generalization of the Kakutani fixed point theorem with application to Nash equilibrium points[J]. The National Academy of Sciences, 1952, 3(1): 170－174.

[12] DEBREU, Gerard. A social equilibrium existence theorem[J]. The National Academy of Sciences, 1952, 38(10): 886－893.

[13] YU Z, WU W, YANG H. The existence of Nash equilibria of population games[J]. Guizhou Science, 2013, 31(2): 28－30.（于曾梅, 武文俊, 杨辉. 群体博弈 Nash 平衡的存在性[J]. 贵州科学, 2013, 31(2): 28－30.）

[14] WU D, LING Y, ZHU H, *et al.* Research on pricing game strategy for load-balancing in VANET[J]. The Journal of China Universities of Posts and Telecommunications, 2013, 20(1): 73－78.

[15] FUDENBERG D, NOWAK M A, TAYLOR C, *et al.* Evolutionary game dynamics in finite populations with strong selection and weak mutation[J]. Theoretical Population Biology, 2006, 70(3): 352－363.