



基于再励学习与遗传算法的 交通信号自组织控制¹⁾

杨煜普 欧海涛

(上海交通大学自动化系 上海 200030)

(E-mail: ypyang@public1.sta.net.cn)

摘 要 提出一种基于再励学习和遗传算法的交通信号自组织控制方法. 再励学习针对每一个道路交叉口交通流的优化, 修正每个信号灯周期的绿信比. 遗传算法则产生局部学习过程的全局优化标准, 修正信号灯周期的大小. 这种方法将局部优化和全局优化统一起来, 克服了现有的控制方法需要大量数据传输通讯、准确的交通模型等缺陷.

关键词 交通系统, 信号灯控制, 再励学习, 遗传算法

中图分类号 TP273

SELF-ORGANIZED CONTROL OF TRAFFIC SIGNALS BASED ON REINFORCEMENT LEARNING AND GENETIC ALGORITHM

YANG Yu-Pu OU Hai-Tao

(Department of Automation, Shanghai Jiaotong University, Shanghai 200030)

(E-mail: ypyang@public1.sta.net.cn)

Abstract A combinative algorithm of reinforcement learning and genetic algorithm is proposed in this paper and is applied to self-organized control of the traffic signals. The reinforcement learning focuses on the optimization of intersection's traffic flow which modifies the split of traffic signal cycle, while the genetic algorithm intends to introduce a global optimization criterion to each of the local learning processes which modifies the cycle itself of traffic signals. This approach overcomes the drawbacks in existing control method such as huge data transfer and communication, accurate traffic model and so on.

Key words Traffic system, signal control, reinforcement learning, genetic algorithm

1 引言

城市交通中各交叉口处交通信号灯控制的优化问题是现代交通控制的重要研究领域,

1) 国家“863”项目(2001AA422420-02)资助

收稿日期 2000-01-17 收修改稿日期 2001-04-15

优化的主要问题是针对交通信号灯控制的信号周期、相位差和绿信比这三个参数进行的,优化的主要方法可分为离线节点优化控制和在线节点优化控制两大类^[1,2]. 其中的在线节点优化控制方法能够根据车辆实际情况,做到不同时段绿灯信号的动态分配,实现交通信号灯闭环反馈控制,代表着现代交通优化控制的研究热点和发展方向. 但是,现有的在线节点优化控制方法需要有准确交通模型、大量的现场信号和大量的数据传输通讯,使得实际应用效果受到很大限制. 因为在实际应用过程中要想获得准确的交通模型是一件非常困难的事情;其次,大量交通信号及通讯处理使得中央控制器面临解决大规模的组合优化问题,又为系统的实时性带来很大困难. 特别是对于交通状态发生突变的情况,如交通事故引起的拥塞等,现有的方法就很难理想的完成优化控制任务^[3]. 我们期望获得一种自适应控制器,能在不引起组合爆炸搜索的情况下在线进行全局目标优化. 因此,本文提出一种遗传再励学习的交通信号自组织控制方法,本方法通过再励学习修正每个交叉口信号灯周期的绿信比实现局部交通优化,通过遗传算法优化信号灯周期形成全局优化标准.

2 信号灯控制的遗传再励学习

2.1 局部再励学习

为适应未知的动态交通流量变化,尤其为了快速适应环境变化,信号灯控制器要在没有中央控制器提供控制指令的情况下,根据自身传感器的信息决定采用何种行动,这是一个局部优化控制问题. 本文采用再励学习的方法实现局部优化控制,过程如下

1) 信号灯控制器行动指令:定义 C_i 为第 i 个交叉口, C_i 的所有信号灯都由控制器 B_i 控制. B_i 发出的行动指令为 a_0, a_1 , a_0 表示进行状态转换,如从红灯变为绿灯, a_1 表示保持原来的状态.

2) 控制器的观测状态:假定 B_i 能观测相邻信号灯的状态,定义 P_j 是相邻交叉口 C_j 的一个信号灯的状态,则 B_i 观测到的状态为

$$S_i = \sum_{j \in J_j} P_j \cdot N^j \quad (1)$$

其中 N 是状态数 ($N=2$), J_j 是相邻交叉口的集合.

3) 行动指令 a_j 的评价:若用 f_k^i 表示 C_i 在信号灯第 k 周期时的各方向交通流量总和,则第 k 周期时 B_i 行动指令 a_j 的评价定义为

$$b = f_{k-1}^i - f_k^i \quad (2)$$

如果 $b > 0$, 则状态 S_i 时的 a_j 应得到再励,否则减弱,用这个准则训练控制器.

4) 行动指令 a_j 的再励学习:定义 $p_j(S_i)$ 为控制器 B_i 在状态 S_i 时选取行动 a_j 的概率. 在一个周期开始时,控制器 B_i 观测到状态 S_i , 并根据概率 $P_0(S_i), P_1(S_i)$ 选取行动 a_j , 然后在下一个周期开始时计算评价标志函数

$$e = \text{sgn}(b \geq 0) = \begin{cases} 1, & \text{if true} \\ 0, & \text{if false} \end{cases} \quad (3)$$

如果 $e=0$ 表示 $p_j(S_i)$ 不适合,减少 $p_j(S_i)$ 而增大 $p_{i \neq j}(S_i)$. 对概率 $p_j(S_i)$ 的修正采用 Narendra 的随机再励学习策略^[4], 即对状态 S_i 选取行动的概率修正算法为

$$\text{if } e = 1 \text{ then } p_j(S_i) \leftarrow p_j(S_i) + \alpha(1 - p_j(S_i)),$$

$$\begin{aligned}
 p_{i \neq j}(S_i) &\leftarrow (1 - \alpha) p_j(S_i) \\
 \text{else } p_j(S_i) &\leftarrow (1 - \beta) p_j(S_i) \\
 p_{i \neq j}(S_i) &\leftarrow (1 - \beta) p_j(S_i) + \beta / (N - 1)
 \end{aligned} \tag{4}$$

其中 α, β 是修正因子.

5) 确定控制绿信比 R_g : 如果 $p_j(S_i) > p_{i \neq j}(S_i)$, 则 $p_j(S_i)$ 对应的 a_j 在信号周期开始时执行, 因此得到绿信比

$$R_g = p_j(S_i) / p_{i \neq j}(S_i) \tag{5}$$

2.2 遗传算法全局优化

交叉口控制器的再励学习可使局部交通流量达到最大, 然而局部最优并不能保证全局最优. 很多情况下一个交叉口处的交通流量的优化反而会引起其他交叉口处的交通拥塞, 因此需要一个全局标准来进行协调. 本文通过修正再励学习的系统参数进行全局优化, 采用遗传算法实现对给定参数的寻优选取. 步骤如下

1) 编码机制: 控制器 B_i 的编码参数取为其信号灯周期 U_i , 在 U_i 给定空间内初始随机编码 M 个点: $U_i(j), j=1, \dots, M$, M 是群体大小. $Q(j) = \{\forall_i U_i(j)\}$ 表示一个区域中 N 个交叉口的 $U_i(j)$ 依次排列形成染色体的基因串, N 表示基因串长度.

2) 适应度函数: 遗传算法的目标函数是搜索适应度函数的最大值. 定义适应度函数为

$$F = \sum_{i,k} f_k^i(U_i(j)) \tag{6}$$

其中 $f_k^i(U_i(j))$ 由局部再励学习过程得到.

3) 控制参数: 算法中的控制参数有群体大小 N 、交换率 P_c 、变异率 P_M 和遗传代数 G . 其中变异率采用随时间变化的自适应搜索策略, 即

$$P_M(j) = (P_M^d)^j \times P_M^0, \quad j = 0, 1, \dots, G - 1 \tag{7}$$

其中 P_M^0 为初始变异率, P_M^d 为每一遗传子代变异率的递减速率.

4) 遗传算子: 包括复制算子、交换算子和变异算子. a. 复制算子: 本文中的复制算子采用轮盘赌选择机制, 选择 $Q(j)$ 的机会正比与 $(F_j - F_{\min})$. b. 交换算子: 设一个串的长度为 l , 随机产生一个整数 $j \in [1, l]$, 然后交换这两个后代的第 j 个点(基因)的左边部分. c. 变异算子: 在本文的应用中, 变异过程是把通过变异概率测试的串中某基因 $U_i(j)$ 的值变为某范围内随机选取的一个值.

5) 初始化和结束准则: 从参数空间内随机地选择 M 个点组成初始群体, 当算法已经迭代到了预置的代数, 则停止迭代.

3 仿真实验

仿真实验设计如下: 在一个正方形的 3×3 交通网路中, 车辆在各条道路上按泊松分布到达, 速度为 d_i ; 定义方向

$$d_i = \{Right, Forward, Left\}, \quad i = 1, 2, 3 \tag{8}$$

车辆经过一个交叉口变向概率为 p_{a_i} , 我们取为 $p_{a_i} = \{0.33, 0.33, 0.33\}$. 当信号灯从红变为绿时, 等待的车辆要经过 t_{accel} 时间才能移动, 即从 0 加速到速度 d 所需时间. 交叉口的长度为 $12\text{m} \times 12\text{m}$ 的正方形, 车辆队列的密度为 $8\text{m}/\text{辆}$, 仿真具体参数如表 1 所示.

表 1 计算机仿真的具体参数

参数	取值	参数	取值
交叉口数目	3×3=9	车辆到达的平均间隔	4s(未饱和的情况下)
交叉口之间的距离	150m	再励学习中的参数	$\alpha=0.2, \beta=0.01$
信号灯周期 $U_i(j)$	学习控制和随机控制为[10s,24s]; 常规控制为[20s,48s]	每个实验的时间长度	40min
相位差	0s	遗传算法中的群体大小	6
绿信比	由交叉口控制器决定	初始变异率	0.2
加速时间 t_{accel}	4s	变异率的递减速率	0.99
道路中的最大车辆数目	10,30,or 50	交换率	0.5

仿真实验采用以下三种信号灯方式进行对比研究

1) 随机信号灯方式:每个信号灯周期开始时,信号灯状态改变的概率为 50%,绿信比为 1/2;

2) 常规信号灯方式:每个信号灯周期开始时都改变信号灯状态,绿信比为 1/2;

3) 再励学习方式:由再励学习算法确定状态转换和绿信比.

本文对这三种情况的信号灯周期时间均采用遗传算法在全局范围内寻优,遗传算法中每个 $U_i(j)$ 的实验时间长度为 40 分钟. 设置最大存在车辆数 30,50,分别代表交通流量“一般”和“拥挤”的情况. 图 1 和图 2 给出了这两种情况下网路中交通流量总和随遗传算法代数的变化过程,其中的纵轴表示在 40 分钟内网路中所有 9 个交叉口的交通流量总和,横轴表示遗传算法的代数.

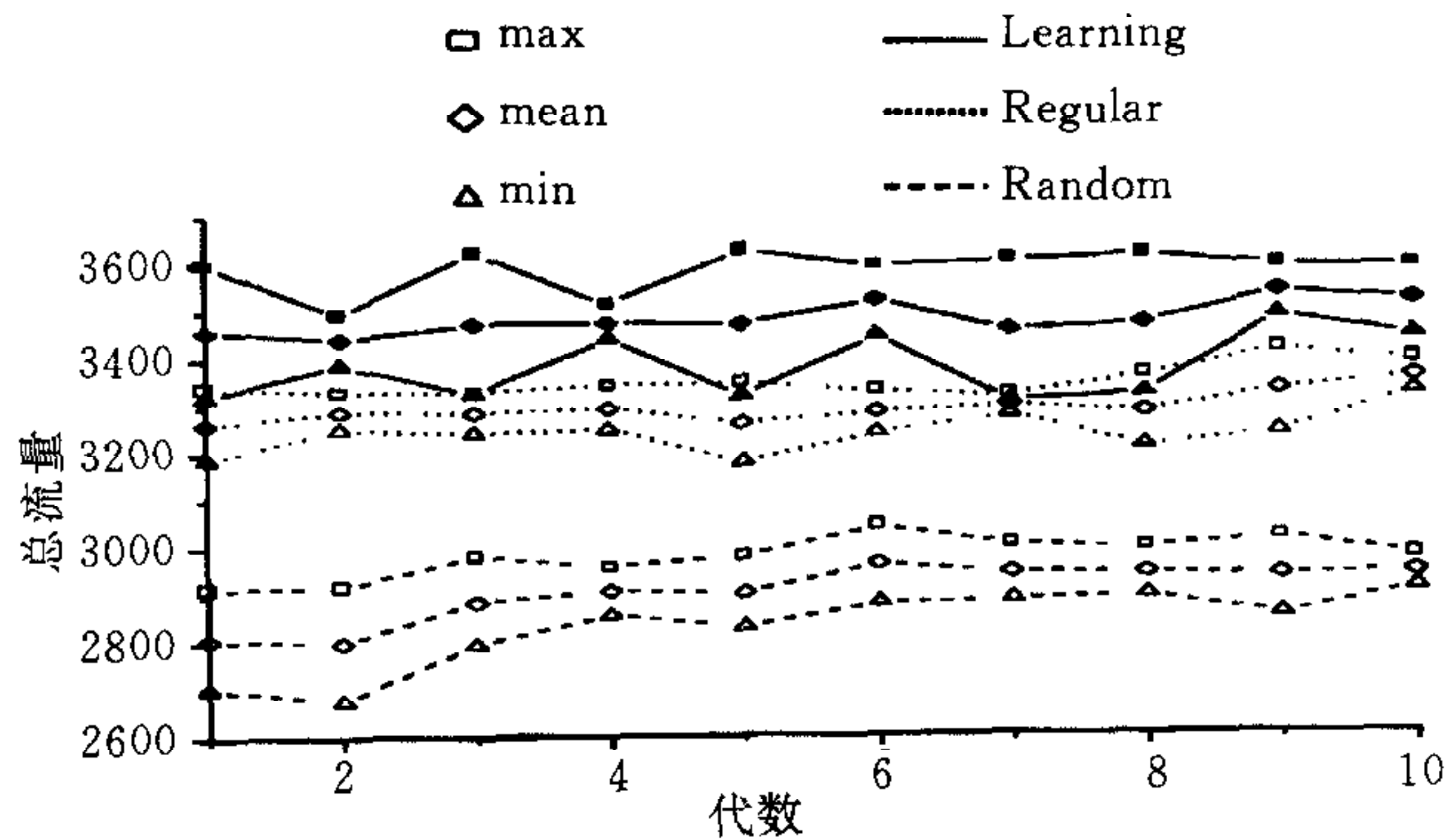


图 1 交通流量一般情况下的仿真结果

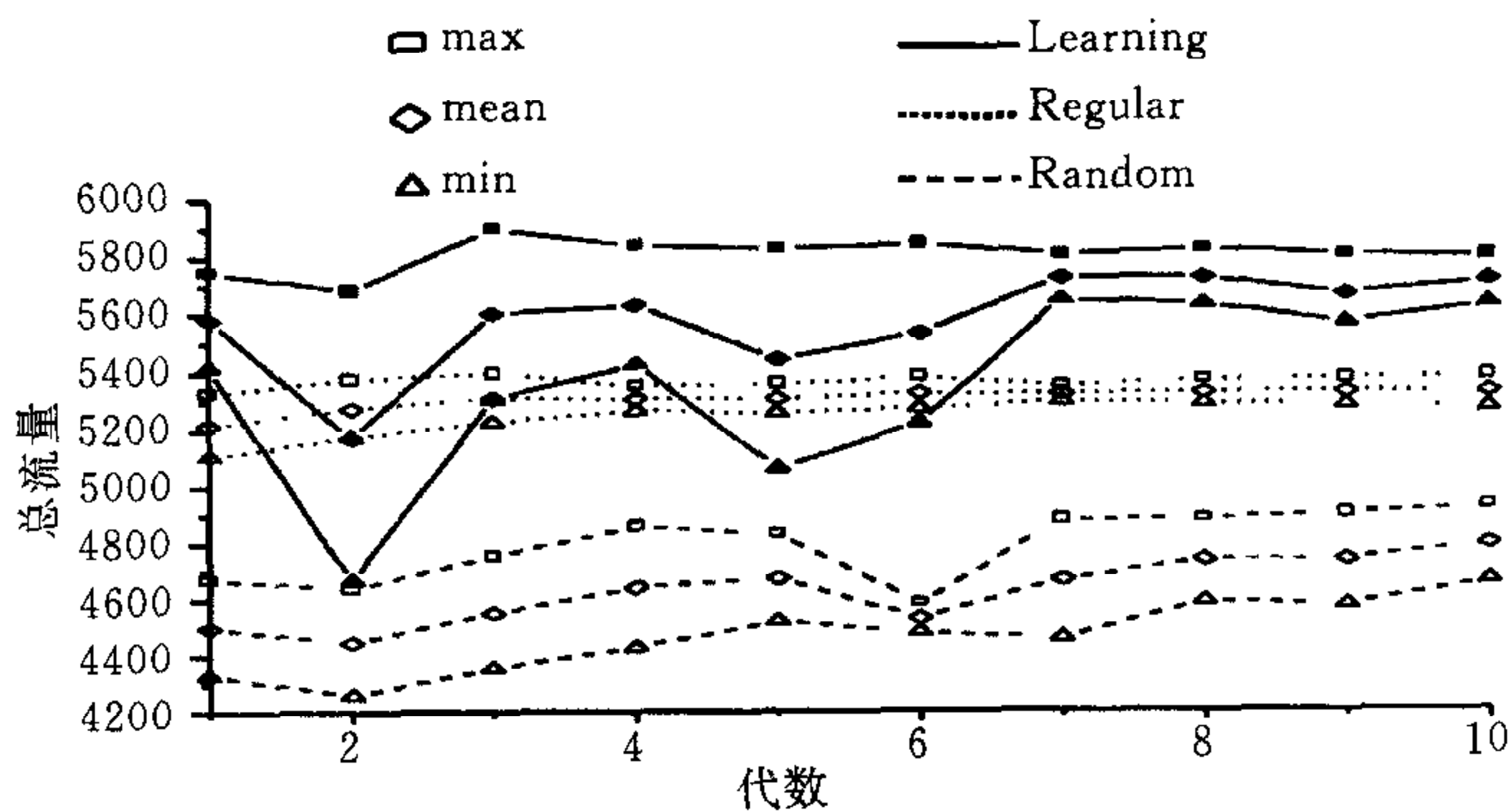


图 2 交通流量拥挤情况下的仿真结果

图中显示通过遗传算法对全局信号灯周期的优化,使这三种信号灯控制方法都能增加平均交通流量,而再励学习方法比其他两种方法能得到更大的交通流量.另外,图 2 的再励学习方式在搜索开始时(从第 3 代到第 6 代)适应度值变化较大,而在第 8 代之后所有的群体基本上都收敛到相同的高适应度值.仿真表明,尽管每次实验车辆出现的模式均不相同,但都能收敛到高适应度值,说明应用再励学习遗传算法可以在动态变化的交通环境下找到鲁棒性较好的控制参数,在交通流量拥塞的状况下能得到更好的控制效果.

4 结论

本文提出一种遗传再励学习算法对交通信号进行优化控制的方法,克服了现有的控制方法需要的大量数据传输通讯、实时性较差、需要准确的交通模型等缺陷.本文的方法将局部优化和全局优化统一起来,并通过计算机仿真实验表明了本方法的有效性,为实现更为有效的交通网络控制提供了新的方法和思路.

参 考 文 献

- 1 王亦兵,韩曾晋,贺国光. 城市高速公路交通控制综述. 自动化学报,1998,24(4):484~496
- 2 Yang H, Yanger S. Traffic assignment and signal control saturated road networks. *Transportation Research Part B*, 1995, 29(2):125~139
- 3 Roozmond D A, P van der Veer. Intelligent transport systems: Autonomous urban traffic control. In: Proceedings of the 4th World Congress on Intelligent Transport Systems, Berlin, 1997. 21~24
- 4 Narendra K, Thathachar M A. *Learning Automata*. Boston: Addison Wesley, 1994. 62~70

杨煜普 博士、教授. 研究兴趣为智能控制理论及应用以及多智能体协调方法研究.

欧海涛 博士研究生. 研究方向为智能控制理论及应用.