

优化 BP_AdaBoost 算法及其交通事件检测

刘庆华^{1,2}, 丁文涛², 涂娟娟², 方守恩¹

(1. 同济大学 交通运输工程学院, 上海 200092; 2. 江苏科技大学 计算机科学与工程学院, 江苏 镇江 212003)

摘要: 为了及时检测出高速公路上发生的交通事件,减少由于交通事件带来的损失,提出了一种基于遗传优化的 BP_AdaBoost 算法用于交通事件检测. 提取高速公路上下游的车流量、车速与占有率作为 BP(back propagation)神经网络的输入值,利用遗传算法全局搜索的性能优化 BP 神经网络初始连接权值和输出阈值,再通过多个新的 BP 神经网络弱分类器构建 AdaBoost 强分类器,设计基于遗传算法优化 BP_AdaBoost 算法的交通事件分类器. 以在东京高速公路采集的真实数据进行性能验证,试验结果表明,该算法可以提高 BP 弱分类器的性能,检测率达到 97%,误报率降至 3.34%,适用于高速公路交通事件的检测.

关键词: 交通事件; 遗传算法; 神经网络; BP_AdaBoost 算法; 东京高速公路

中图分类号: U491.3

文献标志码: A

Improved BP_AdaBoost Algorithm and its Application in Traffic Incident Detection

LIU Qinghua^{1,2}, DING Wentao², TU Juanjuan², FANG Shouen¹

(1. School of Transportation Engineering, Tongji University, Shanghai 201804, China; 2. College of Computer Science and Engineering, Jiangsu University of Science and Technology, Zhenjiang 212003, China)

Abstract: In order to detect the traffic incidents occurred on highway and reduce the loss brought by traffic incident, this paper presents an improved BP_AdaBoost algorithm based on genetic algorithm for traffic incident detection. The inputs of BP (Back Propagation) neural network value are vehicle quantity, velocity and occupancy in upstream and downstream of highway. Genetic algorithm is used for each BP neural network classification model for optimizing weights and thresholds due to its performance of global searching. The

optimized BP neural network model is applied as a new weak classifier, then through the AdaBoost algorithm, many of these new weak classifier is composed as strong classifier model. This improved algorithm is validated with real data from Tokyo expressway ultra-sonic sensors. The experimental results show that the algorithm can improve the performance of BP weak classifier. The detection rate of improved BP_AdaBoost algorithm is up to 97%, and false alarm rate is lower to 3.34%. Experiment indicate that the algorithm is suitable for detecting highway traffic incidents.

Key words: traffic incident; genetic algorithms; neural network; BP_AdaBoost algorithm; Tokyo expressway

交通事件是指非周期发生的且使某段道路通行能力减弱的事件^[1],包括故障停车、货物掉落、交通事故、道路修护和天气影响等. 由于高速公路车辆运行速度较快,而事件发生的时间和地点是随机的,无法预知,故高速公路发生交通事件时,如果不及时排除,容易造成交通拥堵、财产损失和人身伤害. 因此,对于高速公路交通事件的有效检测对保证交通安全、提高运输效率具有重大的经济效益和社会效益,同时,也是一个国家的智能交通系统(intelligent transport system, ITS)的一个重要研究课题.

迄今为止,针对高速公路交通流的不确定性和非线性,国内外已经开发了多种交通事件自动检测算法(automatic incident detection, AID),如国外的 Cheu 等^[2]将支持向量机模型应用于高速公路事件检测中,进行了建模和仿真研究; Jeong 等^[3]运用离散小波变换对交通流数据进行处理,与设定的阈值进行比较,实现交通事件的检测;在国内,张轮等^[4]设计朴素贝叶斯分类的高速公路交通事件检测算法

收稿日期: 2015-01-29

基金项目: 国家“八六三”高技术研究发展计划(2013AA12A206); 国家自然科学基金(51008143); 江苏省高校自然科学基金(14KJD520002)

第一作者: 刘庆华(1977—),男,副教授,工学博士,主要研究方向为智能交通与道路安全. E-mail: giant_liu@163.com

通讯作者: 方守恩(1961—),男,教授,博士生导师,工学博士,主要研究方向为道路交通安全、道路规划与计算机辅助设计等.

E-mail: fangsek@tongji.edu.cn

并进行了仿真试验;牛世峰等^[5]提出纵向时间序列算法,以提高交通事件检测算法的效率.

目前,BP(back propagation)神经网络在交通事件检测系统中也广泛应用.童飞^[6]运用 BP 神经网络实现水上交通事件的检测;陈君等^[7]构建 BP 神经网络评价模型运用于高速公路交通安全诊断;Yu 等^[8]使用环路检测数据和浮动车数据通过 BP 神经网络以实现交通事件的检测. BP 神经网络具有超强适应能力和训练能力,但存在的缺点是权值和阈值的合理化设置较难,训练和记忆具有不稳定性.

针对 BP 神经网络的性能受到权值和阈值影响的缺点,遗传算法优化 BP 神经网络算法被提出,并已在图像分类^[9]、参数优化^[10]和故障诊断^[11]成熟运用.本文引入遗传算法^[12]优化 BP 神经网络的初始连接权值和输出阈值,通过该方法得到最佳权值和阈值.为了加强分类效果,再通过多个优化后的 BP 神经网络弱分类器组建成 AdaBoost 强分类器,设计基于遗传算法优化 BP_AdaBoost 算法的交通事件分类器.

1 优化的 BP_AdaBoost 算法分析

1.1 传统的 BP 神经网络

BP 神经网络是一种单向传播的多层前馈型神经网络,它具有处理非线性复杂系统问题的能力^[13]. BP 网络由输入层、1 个或多个隐层(中间层)和输出层组成,图 1 为隐层 1 层的 BP 网络模型结构图.

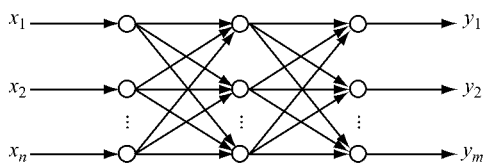


图 1 隐层 1 层的 BP 网络模型结构

Fig.1 BP structure with one hidden layer

当把训练样本提供给网络后,神经元的激活值从输入层经过隐层向输出层传播,在输出层的神经元获得网络的输出.在此训练过程中,通过修正网络的连接权值来减少目标输出与实际的误差,最终达到期望的结果.标准的 BP 网络算法采用均方误差作为表示网络性能的函数,网络的学习过程就是使式(1)减小到可接受的程度.

$$M_{SE} = \frac{1}{2} \sum_{i=1}^m (t_i - y_i)^2 \quad (1)$$

式中: M_{SE} 为网络均方误差; I 为节点号; m 为输出节

点数; t_i 为网络的期望输出值; y_i 为网络的实际输出值.

1.2 遗传算法优化 BP 神经网络

BP 神经网络的权值和阈值一般是通过随机初始化为 $[-0.5, 0.5]$ 区间的随机数,这个初始化参数对网络性能影响很大,但是又无法准确获得,因此,采用遗传算法优化初始权值和阈值.遗传算法优化 BP 神经网络的要素包括种群初始化、适应度函数、选择算子、交叉算子和变异算子. BP 神经网络的结构已知,权值和阈值的个数就已知了,种群中的每一个个体都包含了一个网络所有权值和阈值,个体通过适应度函数计算个体适应度值,再通过选择、交叉和变异找到最优适应度对应的个体,接着,把最优个体赋值给网络的权值和阈值.遗传算法优化 BP 神经网络算法过程如下.

(1) 本文个体编码方法为实数编码,每个个体均为一个实数串,个体包含了 BP 神经网络全部权值和阈值,在网络结构已知的情况下,就可以构成一个结构、权值、阈值确定的神经网络.

(2) 根据个体得到 BP 神经网络的初始权值和阈值,用训练样本训练 BP 神经网络后预测系统输出,把预测输出和期望输出之间的误差绝对值和作为个体适应度值 F ,计算公式为

$$F = Q \left(\sum_{i=1}^n |y_i - o_i| \right) \quad (2)$$

式中: n 为网络输出节点数; y_i 为 BP 神经网络第 i 个节点的期望输出; o_i 为第 i 个节点的预测输出; Q 为系数.

(3) 本文中选择操作的实现方法使用轮盘赌法,即基于适应度比例的选择策略,每个节点 i 的选择概率 p_i 为

$$f_i = R/F_i \quad (3)$$

$$p_i = f_i / \sum_{j=1}^N f_j \quad (4)$$

式中: F_i 为节点 i 的适应度值,由于适应度值越小越好,所以在个体选择前对适应度值求倒数; R 为系数; N 为种群个体数目.

(4) 交叉操作方法采用实数交叉法,第 s 个染色体 a_s 和第 l 个染色体 a_l 在 j 位的交叉操作方法为

$$\left. \begin{aligned} a_{sj} &= a_{sj}(1-b) + a_{lj}b \\ a_{lj} &= a_{lj}(1-b) + a_{sj}b \end{aligned} \right\} \quad (5)$$

式中, b 是 $[0, 1]$ 间的随机数.

(5) 选取第 i 个个体的第 j 个基因 a_{ij} 进行变异,变异操作方法如下:

$$a_{ij} = \begin{cases} a_{ij} + (a_{ij} - a_{\max})f(g) & r \geq 0.5 \\ a_{ij} + (a_{\min} - a_{ij})f(g) & r < 0.5 \end{cases} \quad (6)$$

$$f(g) = r_2(1 - g/G_{\max}) \quad (7)$$

式中: a_{\max} 为基因 a_{ij} 的上界; a_{\min} 为基因 a_{ij} 的下界; r_2 为随机数; g 为当前迭代次数; G_{\max} 是最大进化次数; r 为随机数, $r \in [0, 1]$.

1.3 改进的 BP_AdaBoost 算法

AdaBoost 算法的思想是合并多个“弱”分类器的输出以产生误差小的期望结果. 本文为了增强分类效果,把通过遗传算法优化之后的 BP 神经网络作为新的弱分类器,通过 AdaBoost 算法构建成 BP_AdaBoost 强分类器. 首先,从样本空间中随机选取 L 组交通流数据作为训练样本,初始化训练样本的权重,按下式计算:

$$D_i(r) = 1/L(r \in 1, 2, \dots, L) \quad (8)$$

再用训练样本数据训练被遗传算法优化之后的 BP 神经网络,当训练第 $k(k \in 1, 2, \dots, M)$ 个 BP 神经网络弱分类器时,得到 n 组数据的预测误差和为

$$e_k = \sum [g_k(r) - y_r](r \in 1, 2, \dots, L) \quad (9)$$

式中: $g_k(r)$ 为网络的期望输出值; y_r 为网络的实际输出值.

进而根据 n 组训练样本的预测误差和 e_k 计算第 $k(k \in 1, 2, \dots, M)$ 个 BP 神经网络弱分类器的权重如下:

$$a_k = \frac{1}{2} \ln\left(\frac{1 - e_k}{e_k}\right) \quad (10)$$

接着,根据权重 a_k ,调整下一轮训练样本的权重,其公式为

$$D_{k+1}(r) = \frac{D_k(r)}{B_k} \exp[-a_k g_k(r) y_i](r \in 1, 2, \dots, L) \quad (11)$$

式中: B_k 为归一化因子; $D_k(i)$ 为经过 $k-1$ 次训练调整后的权重.

最后,建立强分类函数. 训练 T 次后,得到由 T 组弱分类函数组合而成的强分类函数,计算式为

$$h(x) = \sum_{k=1}^T a_k f_k(x) \quad (12)$$

式中: $f_k(x)$ 为第 k 个弱分类函数; x 为测试数据.

改进后算法的流程图如图 2 所示,算法采用在东京高速公路采集的真实数据来进行验证试验,算法所得的结果通过 MATLAB 软件仿真得到.

2 交通事件检测试验

2.1 参数选择

本文的交通流数据来源于如图 3 所示的日本首

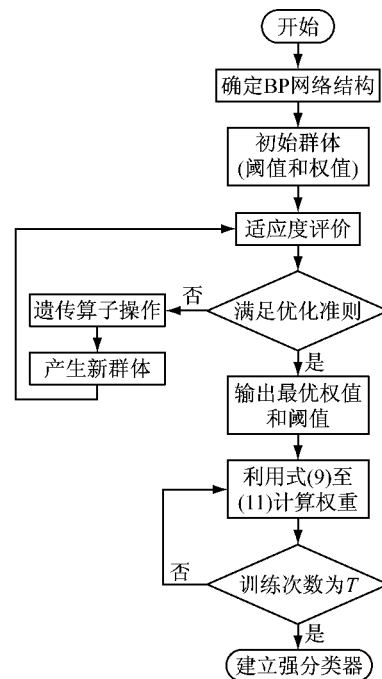


图 2 遗传算法优化 BP_AdaBoost 算法的算法流程
Fig.2 The algorithm progress of genetic algorithm optimized BP_AdaBoost algorithm

都东京高速公路 3 号涩谷线,第 1 车道上检测器在 42 号与 43 号之间的路段,路段长 1 000 m,数据包括车流量(Q)、车速(V)和占有率(O_{cc}),采样周期为 1 min,两检测器之间的距离为 1 km,时间是 2010 年 12 月 1 日至 12 月 31 日.

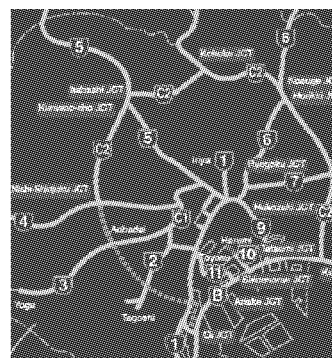


图 3 东京首都高速公路网络

Fig.3 The Tokyo metropolitan expressway network

本文对其数据进行整理分析,分别对事件发生前后上下游的 Q, V, O_{cc} 进行试验分析. 如图 4a 所示,事件发生在 80 min 时刻左右,在 80 min 之前,由于上下游未发生交通拥堵,交通保持畅通,车速保持在 $70 \sim 90 \text{ km} \cdot \text{h}^{-1}$ 之间波动,上下游处于相对稳定状态. 而在 80 min 时刻以后,由于交通事件的发送,上游车辆产生交通拥挤,车辆运行受阻,故车速迅速减小到 $15 \text{ km} \cdot \text{h}^{-1}$ 左右;下游车辆由于未受到交通

事件的干扰,故车速基本维持稳定变化.因此,上下游车速有较明显的变化差异.

交通事件发生前,由于交通处于正常运行状态,上下游占有率均在 5%左右,故检测器监测到的上下游占有率处于动态平衡状态.而在 80 min 时刻发生了交通事件,导致上游交通发生堵塞,故占有率骤增到 35%左右;由于下游未受到交通事件的影响,同时来源于上游的车辆数急剧减少,因此,下游占有率出

现一定程度的降低.具体直观显示如图 4b 所示.

如图 4c 所示,事件发生前,上下游的车流量分布处于相对平衡状态.而事件发生后,上下游车流量有下降趋势,并且两者的变化趋势大体一致.

通过对图 4a 至图 4c 的分析,事件发生前, Q , V , O_{cc} 都会处于动态平衡状态;而事件发生以后,上下游的这 3 个交通流参数均会发生明显的变化.

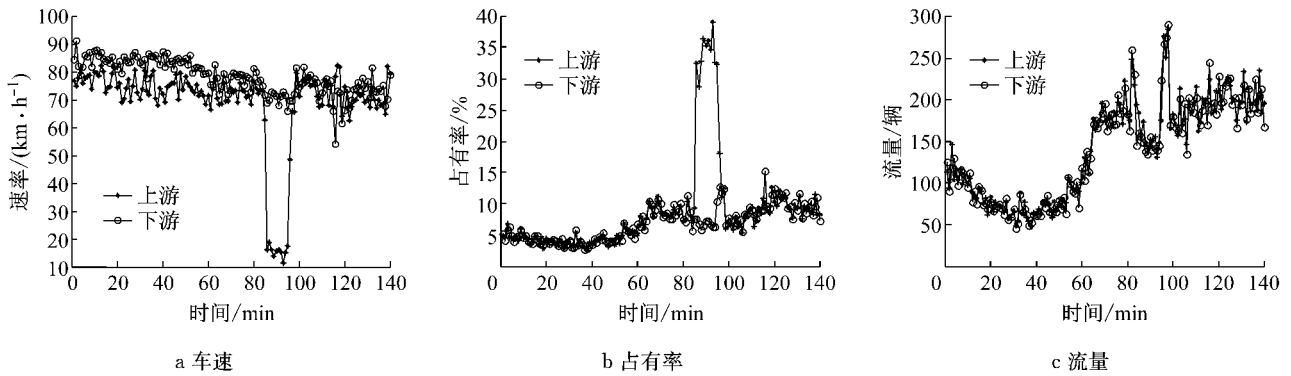


图 4 车速、占有率及流量变化

Fig. 4 Speed of vehicle and occupancy rate and traffic volume

2.2 模型建立

本文选取其中有事件发生和无事件发生的共 2 000 组数据(包含事件数据 75 例),其中 1 800 数据作为训练样本数据,200 组数据作为测试数据.部分样本数据如表 1 所示.

表 1 部分交通流训练样本数据

Tab.1 Part of the traffic flow sampled data

样本序号	上游			下游		
	Q/辆	V/(km·h ⁻¹)	O _{cc} /%	Q/辆	V/(km·h ⁻¹)	O _{cc} /%
1	158	27.4	19.5	176	15.4	36.7
2	150	15.9	33.2	161	12.9	39.9
3	146	13.0	37.0	160	14.7	30.4
4	131	12.9	36.0	174	13.3	36.8
5	255	69.7	12.0	264	74.3	11.2
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1 799	199	73.1	9.2	185	75.8	7.9
1 800	210	70.8	9.9	215	74.4	9.6

由表 1 可知,交通流参数值变化范围较大,不在同一数量级,因此,有必要将原始交通流数据样本进行归一化处理,从而减小数值差异带来的影响.

本文构建的 BP 神经网络有 6 个输入、2 个输出,6 个输入分别为上游的 Q , V , O_{cc} ,下游的 Q , V , O_{cc} ,2 个输出代表事件发生情况,输出为 {1,0} 表示有事件发生,输出为 {0,1} 表示无事件发生.所以,设置 BP 神经网络的网络结构为 6-13-2,即输入层有 6 个结点、隐含层有 13 个结点、输出层有 2 个结点.整

个 BP 神经网络结构一共有 $6 \times 13 + 13 \times 2 = 104$ 个权值及 $13 + 2 = 15$ 个阈值,由此可得,遗传算法优化参数的个数为 $104 + 15 = 119$.

BP 网络隐层神经元传递函数采用 tansig 函数,输出层采用 purelin 函数.训练次数为 50 步,训练目标为 0.001.

在遗传算法优化 BP 神经网络的程序中,遗传算法的参数设置如下:进化代数为 10 次,种群规模为 10,交叉概率为 0.4,变异概率为 0.2.

在优化的 BP_AdaBoost 算法中,采用 10 个 BP 弱分类器组成的强分类器对数据样本进行检测,其中误差阈值设置为 0.1.

2.3 结果分析

在本文选取的 100 组测试数据中,结果只有 3 组与期望值不一样,其余 97 组与期望值输出结果相同,说明本文提出的优化的 BP_AdaBoost 算法的检测率高达 97.0%.表 2 是经过同样的训练样本训练之后 BP_AdaBoost 算法优化前后的结果比较.

由表 2 可见,BP 神经网络算法的检测率只有 78.0%,BP_AdaBoost 算法的检测率为 89.8%,而经过遗传算法优化的 BP_AdaBoost 算法的检测率高达 97.0%;在误报率上,3 种算法的差距也相当明显.此外,将本文提出的方法与 LVQ(learning vector quantization)神经网络^[14]和小波分析^[15]进行对比试

验,对比结果如表 3 所示.表 3 可见,本文提出的方法比其他 2 种方法检测率高、误报率低,说明本文方法较好.

表 2 优化前后的 BP_AdaBoost 算法试验结果比较

Tab. 2 Comparison of BP_AdaBoost algorithm experimental results before and after optimization

算法名称	DR/%	FAR/%	MTD/%
BP 神经网络算法	78.0	6.41	3.12
BP_AdaBoost 算法	89.8	5.25	4.31
本文方法	97.0	3.34	4.22

表 3 3 种不同交通事件检测算法性能比较

Tab. 3 Three different traffic incident detection algorithm performance comparison

算法名称	DR/%	FAR/%	MTD/%
LVQ 神经网络	74.3	5.26	4.59
小波分析	90.5	4.12	3.97
本文方法	97.0	3.34	4.22

3 结论

提出了优化的 BP_AdaBoost 算法的高速公路事件检测方法,根据遗传算法具有快速搜索最优解的特点,来优化 BP 神经网络初始连接权值和输出阈值,用样本数据训练新的 BP 神经网络,再通过多个新的 BP 神经网络弱分类器构建成 AdaBoost 强分类器,并采用东京高速公路的真实交通流数据进行性能验证.试验结果表明,采用遗传算法优化 BP 神经网络,减小了权值和阈值不确定性对 BP 神经网络性能的影响.与改进后的算法与原算法相比提高了检测率,降低了误报率,具有更好的检测性能,适用于高速公路的交通事件检测.

参考文献:

- [1] Zhang K, Taylor M A P. Towards universal freeway incident detection algorithms [J]. *Transportation Research Part C: Emerging Technologies*, 2006, 14(2): 68.
- [2] Cheu R L, Srinivasan D, Teh E T. Support vector machine models for freeway incident detection [C] // *Intelligent Transportation Systems*, 2003. Singapore: IEEE, 2003: 238-243.
- [3] Jeong Y S, Castro-Neto M, Jeong M K, *et al.* A wavelet-based freeway incident detection algorithm with adapting threshold parameters [J]. *Transportation Research Part C: Emerging Technologies*, 2011, 19(1): 1.
- [4] 张轮, 杨文臣, 刘拓, 等. 基于朴素贝叶斯分类的高速公路交通事件检测 [J]. *同济大学学报: 自然科学版*, 2014, 42(4): 558.
- [5] 牛世峰, 姜桂艳, 李红伟, 等. 基于纵向时间序列的快速路交通事件检测算法 [J]. *哈尔滨工业大学学报*, 2011, 43(2): 144.
- [6] 童飞. 基于 BP 神经网络的水上交通事故预测及 MATLAB 实现 [D]. 武汉: 武汉理工大学, 2005.
- [7] 陈君, 李聪颖, 丁光明. 基于 BP 神经网络的高速公路交通安全评价 [J]. *同济大学学报: 自然科学版*, 2008, 36(7): 927.
- [8] Yu L, Yu L, Wang J, *et al.* Back-propagation neural network for traffic incident detection based on fusion of loop detector and probe vehicle data [C] // *Natural Computation*, 2008. ICNC '08. Fourth International Conference on Natural Computation. Jinan: IEEE, 2008: 116-120.
- [9] Liu Z, Liu A, Wang C, *et al.* Evolving neural network using real coded genetic algorithm (GA) for multispectral image classification [J]. *Future Generation Computer Systems*, 2004, 20: 1119.
- [10] Shen C, Wang L, Li Q. Optimization of injection molding process parameters using combination of artificial neural network and genetic algorithm method [J]. *Journal of Materials Processing Technology*, 2007, 183: 412.
- [11] 段侯峰. 基于遗传算法优化 BP 神经网络的变压器故障诊断 [D]. 北京: 北京交通大学, 2008.
- [12] Duan Houfeng. Genetic algorithm optimization of transformer fault diagnosis based on BP neural network [D]. Beijing: Beijing Jiaotong University transformer fault diagnosis, 2008.
- [13] Ding S, Su C, Yu J. An optimizing BP neural network algorithm based on genetic algorithm [J]. *Artificial Intelligence Review*, 2011, 36: 153.
- [14] Li Z, Lei Q, Kouying X, *et al.* A novel BP neural network model for traffic prediction of next generation network [C] // *Natural Computation*, ICNC 09, International Conference on Natural Computation. Tianjin: IEEE, 2009: 32-38.
- [15] 朱红斌. LVQ 神经网络在交通事件检测中的应用 [J]. *计算机工程与应用*, 2009, 44(34): 213.
- [16] ZHU Hongbin. Application of neural network in traffic incident detection [J]. *Computer Engineering and Application*, 2009, 44(34): 213.
- [17] 彭宇. 基于小波分析的高速公路事件检测算法 [J]. *湖南交通科技*, 2011, 37(3): 123.
- [18] PENG Yu. Freeway incident detection method based on wavelet analysis [J]. *Hunan Traffic Science and Technology*, 2011, 37(3): 123.