

双曲正切函数在事故黑点鉴别中的应用

杨 轸, 唐 莹, 方守恩

(同济大学 交通运输工程学院, 201804 上海, yangzhen2276@263.net)

摘 要: 为改善双指数函数 $y = a \times e^{bx} + c \times e^{dx}$ 在累计频率曲线拟合和事故黑点鉴别时存在的问题, 提出用双曲正切函数 $y = a + b \times \tanh(cx + cd)$ 作为替代, 以提高黑点鉴别效果. 通过仿真和实际数据验证对2种拟合算法进行对比分析. 结果表明: 累计频率曲线采用双曲正切函数进行拟合, 其拟合效果好, 事故黑点鉴别检出率较高, 黑点鉴别的结果比较稳定, 对不同事故统计年限的事故黑点鉴别有很好的适应性. 此外, 一定事故率水平下由累计交通量决定的事故次数对事故黑点的误检率有非常重要的影响. 双曲正切函数可替代双指数函数进行事故黑点鉴别, 同时为将误检率控制在20%左右, 建议待鉴别道路的平均每公里事故次数不小于4次.

关键词: 事故黑点; 鉴别; 双曲正切; 误检率

中图分类号: U491.3

文献标志码: A

文章编号: 0367-6234(2011)10-0143-06

Application of hyperbolic tangent function for accident prone location identification

YANG Zhen, TANG Ying, FANG Shou-en

(School of Transportation Engineering, Tongji University, 201804 Shanghai, China, yangzhen2276@263.net)

Abstract: In order to improve the ability of the accident prone location identification, the hyperbolic tangent $y = a + b \times \tanh(cx + cd)$ is proposed as a substitute fitting function of the double exponential function $y = a \times e^{bx} + c \times e^{dx}$ for fitting scatter diagram of accident cumulative frequency. The comparative analysis of the two fitting functions was conducted by simulation analysis and actual data confirmation. The results indicated that the hyperbolic tangent function has higher correlation coefficient in curve fitting, higher detection rate in accident prone location identification, more stable results and better adaptability for the cumulative traffic volume. More significant finding is proposed that the accident number, which decided by cumulative traffic volume in certain accident rate, has very important influence on misdetection rate of accident prone location identification. The hyperbolic tangent can be used for fitting scatter diagram of accident cumulative frequency instead of the double exponential function, and the suggested number of average accident per kilometer should not less than 4 to control the misdetection ratio to about 20%.

Key words: black spot; identification; hyperbolic tangent; misdetection rate

交通事故在道路上是非均匀分布的, 道路路段中具有突出事故频率的事故位置被称为事故黑点. 传统的事故黑点鉴别方法有事故数法、事故率法、传统贝叶斯法等, 近年来, 国内外交通安全领

域学者对事故黑点鉴别作了很多深入的研究, 提出了经验贝叶斯法、质量控制法、累计频率法等^[1-7]. 其中累计频率法被广泛地应用于公路事故黑点的鉴别中, 它特别适应中国目前基础数据相对缺乏, 各地道路安全状况差别大的实际情况^[8].

在累计频率法中, 需要选取合适的公式对累计频率散点图进行拟合, 拟合公式的选取对黑点

收稿日期: 2010-12-03.

基金项目: “十一五”国家科技支撑计划项目(2007BAK35B03);

国家自然科学基金资助项目(70801044).

作者简介: 杨 轸(1974—), 男, 博士, 副教授.

鉴别有一定的影响. 常用的拟合公式有多项式、指数函数和双指数函数等, 其中中华人民共和国公共安全行业标准《道路交通事故多发位置认定方法(征求意见稿)》提出采用双指数函数作为累计频率法的拟合公式. 在实际使用中, 这些拟合公式或多或少都存在一定的缺陷, 多项式拟合单调性很差, 指数函数和双指数函数 $y = a \times e^{bx} + c \times e^{dx}$ 单调性较好, 通常效果也不错, 但当路段事故集中程度很高或者事故统计时间很长时, 曲线拟合的相关性变差, 需要提出一种适应性更好的拟合函数.

此外, 以往的事故黑点鉴别方法中, 尚未有学者对鉴别结果的有效性进行研究, 特别是定量分析不同累计交通量和不同检测方法对事故黑点鉴定的检出率和误检率的影响. 《道路交通事故多发位置认定方法(征求意见稿)》中要求分析所用的事故数据时间为 1~3 a, 但并未就统计年限对检出率和误检率的影响进行详细说明, 并且不同的道路其交通量和事故率差异很大, 用统一的统计时间来限定值得商榷.

由于事故发生具有随机性, 仅从事事故的统计结果不能直接推导出路段事故率的期望值. 为此, 本文提出了基于随机仿真的黑点位置鉴别比较方法, 其思路为: 首先从多年运营的代表性道路的事故资料中分别统计获得各路段单元的事故率; 然后, 根据该事故率组按照事故发生统计规律通过计算机模拟产生不同累计交通量下一系列事故数据, 并根据这些数据采用不同的拟合方法进行黑点鉴别; 最后, 将拟合结果同期望事故黑点域进行比较, 分析不同鉴别方法在曲线拟合相关性, 不同累计交通量下事故黑点检出率、误检率等方面的差异以推荐最佳的鉴别方法及适应条件. 考虑到仿真的随机性, 可以采用多次仿真(一般不少于 10 次)进行综合比较.

1 事故仿真

对于具有事故率为 P_0 (车公里事故率) 的路段, 其发生交通事故的次数符合二项分布^[4], 即

$$P(X = K) = \binom{m}{k} P_0^k (1 - P_0)^{(m-k)}.$$

其中: k 为发生的事故次数; m 为道路上累计作用的交通量(车公里); P_0 为路网或某类道路的平均事故率(车公里事故率).

对于一段较长公路而言, 各路段的道路设计指标和周边环境有所差异, 各路段的事故率也各不相同, 事故率高的路段往往表明该路段存在缺

陷. 因此, 在分析不同路段上一定累计作用交通量的事故发生次数概率时, 应该采用该路段的事故率, 其期望值可以从多年运营的道路事故频数进行估计. 而进行事故次数仿真时, 该路段的事故发生次数符合其对应事故率的贝努利实验.

对于一条 100 km 长的高速公路, 采用传统方法进行分段, 以 1 km 为一个单元, 把公路分为 100 段, 以沪杭高速公路浙江段 2005 年至 2009 年的下行方向事故数据为基础对每个路段取一个合适的事故率, 如图 1 所示. 运用数理统计的原理, 在 Matlab 中运用 binornd 函数对每个路段随机产生一系列的事故数据, 可以把这些事故次数看作是每个路段在对应统计年限内发生的事故次数. 函数中累计交通量的不同代表事故统计年限的变化, 统计年限越长, 对应在路段上累计交通量就越大. 3 类典型统计时间分别定义为统计年限短(累计交通量取 10^6), 统计年限适中(累计交通量取 5×10^6), 统计年限长(累计交通量取 5×10^7), 按照高速公路单向平均日交通量为 2.5×10^4 计算, 分别相当于 1.5 个月, 7 个月, 5.5 a 交通量. 每类统计年限, 随机产生 $10 \times 100 = 1\,000$ 个事故数.

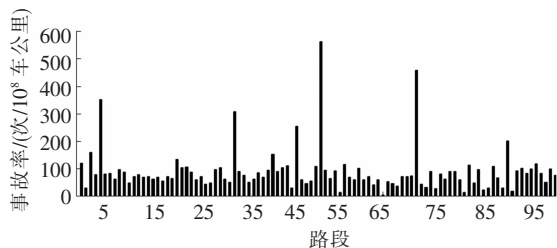


图 1 路段事故率

2 双曲正切函数拟合分析

根据《道路交通事故多发位置认定方法(征求意见稿)》, 对随机仿真产生的事故数据进行分析, 可绘制累计频率散点图. 从散点图的形状方面进行分析, 当事故统计时间较短时, 累计频率散点呈凸形, 这与双指数函数形状比较接近, 而当事故数据统计时间很长时, 累计频率散点呈现 S 型, 而且具备明显的渐进线, 这与双指数函数的形状有一定的差异. 为了完善累计频率法, 可以考虑选取与累计频率散点图变化趋势近似的曲线作为拟合公式. 考察双曲正切函数 $y = a + b \times \tanh(cx + cd)$ 的特点, 通过平移和缩放该曲线, 在 $[0, 1]$ 区间内可以产生不同的形状, 可较好适应不同条件下的累计频率散点拟合. 图 2 为不同统计年限下事故黑点累计频率曲线散点图与双曲线正切拟

合.图 3 为事故统计年限很长时双指数函数与双曲正切函数拟合的对比分析.

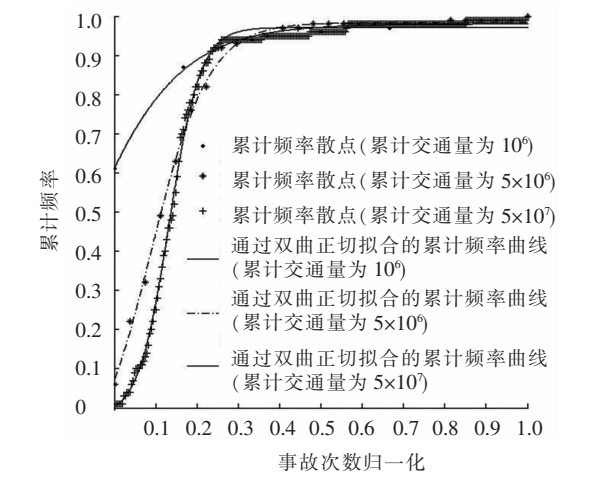


图 2 累计频率曲线散点图与双曲正切拟合

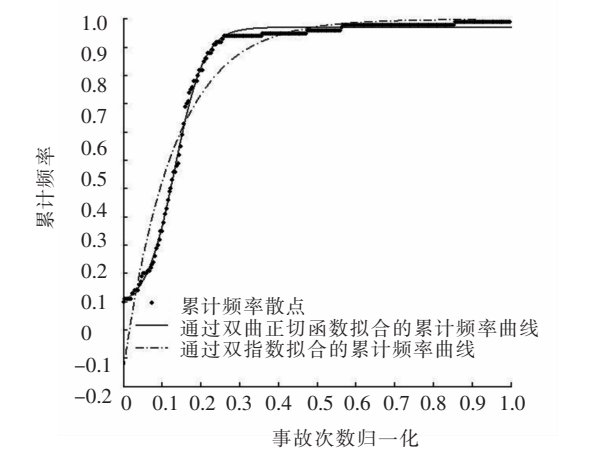


图 3 累计频率拟合曲线对比图

不难发现,双曲正切函数能较好适应不同事故统计年限下的累计频率曲线拟合,尤其对于统计年限较长的情况相对双指数函数来讲具有更好的适应性.

表 1 分别是 3 种不同交通年限下 10 次仿真后对累计频率曲线采用双曲正切函数拟合的参数统计结果.

表 1 双曲正切函数拟合参数统计结果

参数	累计交通量	a	b	c	d
均值	10^6	0.365 2	0.624 9	4.761 8	0.031 3
	5×10^6	0.387 9	0.590 9	8.491 0	-0.075 8
	5×10^7	0.457 5	0.514 8	16.621 9	-0.123 0
标准偏差	10^6	0.108 3	0.108 0	1.153 7	0.054 3
	5×10^6	0.073 4	0.074 0	1.525 9	0.019 7
	5×10^7	0.012 4	0.012 7	1.032 5	0.007 9

分析可知:累计交通量越大(对应事故数据统计年限越长), a 越大, b 越小, c 越大, d 越小;当

累计交通量小时, d 值为正,当累计交通量适中时, d 的符号不确定,当累计交通量大时, d 值为负, d 的正负可以反映事故数据的大概统计年限;统计年限越长,参数的相对标准偏差越小,说明黑点鉴别越稳定.

3 黑点鉴别方法有效性分析

为了分析双指数函数与双曲正切曲线拟合鉴别的差异性,将不同统计年限下的 10 次仿真数据产生的累计频率分别按不同的拟合公式进行拟合,求出拟合曲线上的最小曲率半径的位置作为事故黑点鉴别的临界值,并根据该临界值查找出相应的事故黑点的位置,与期望黑点位置进行比较.

期望黑点域表示真实的黑点位置.理论上,对于当道路上的累计交通量无穷大时,各路段发生的事故次数与其事故率成正比.为了便于分析,将各路段事故率 $\times 10^7$ 作为路段的事故发生次数,并在整数区间进行累计频率计算和黑点鉴别.由于采用双指数函数拟合鉴别出的期望事故黑点数是 6 个,而采用双曲正切函数鉴别出来的结果是 8 个,综合两者,取期望事故黑点个数取 7 个,分别对应的路段是 K51 + 000, K71 + 000, K5 + 000, K33 + 000, K46 + 000, K90 + 000 和 K3 + 000.若鉴别出来的事故位置个数在期望位置域中比重高,则表示检测有效性高,如果检测的位置不在期望域中,则代表误检情况.应该说明的是,误检情况发生不仅受检测方法的影响,同时也受随机数产生的影响,但是从统计的角度对比分析,仍能准确评价检测方法的有效性.

表 2 为双指数函数和双曲正切函数拟合的鉴别结果.对上述事故黑点鉴别结果提取出检出率和误检率,并作进一步的统计分析,如表 3 所示.

对比不同累计交通量下事故黑点鉴别结果可以发现:随着累计交通量的增长,2 种方法对事故黑点的检出率均有不同程度的增加,但增加并不显著;而交通量对误检率的影响非常显著,在累计交通量小时,事故黑点的误检率非常高,甚至大于 70%.这是因为尽管各路段的事故率有所差异,但交通事故的真实发生具有随机性,当事故累计交通量不大时,路段发生的事故次数较少,该随机数的方差与期望值的比值较大,故而各路段真实事故率表现不充分,误检率较高.当累计作用交通量增大,对应的事故次数增加,该随机数的方差与期望值的比值降低,各路段真实事故率表现充分,误检率迅速降低.由于累计频率曲线法是以较长范

围路段实际发生的事故为统计依据,是一种事后检测方法而非预测法,因此在实际操作中若有控制误检率,应给出相应的事故次数. 根据以实际事故数据为基础的仿真分析,若以误检率 20% 左右控制,相当于 5×10^6 累计交通量下产生的事故次数,对应的全路段的平均每公里的发生的事故次数约为 4 ~ 5 次,这对沪杭高速公路来说相当于 7 个月的统计时间. 对于交通量小或者平均事故率很低的道路,为获得较低的误检率,必须有较长的统计年限的事故数据为基础,使各路段的事故率表达充分. 对比 2 种函数的累计频率曲线拟合黑点鉴别结果可以得出:

1) 在较低累计交通量下 2 种函数拟合的效果差别性不大,相关系数都较高. 随着交通量的增

大,双指数函数的拟合相关性有所降低,尤其是在曲率最大位置(黑点判别临界值)与实际情况相差较大(见图 3),而双曲正切函数仍然保持很高的相关性,曲线拟合效果也好.

2) 双曲正切函数在事故黑点的检出率上稍高于双指数函数,在误检率上略差于后者. 这是因为两者各自的期望事故黑点数有所差异,前者高于后者,而为了比较本文采用了两者的平均值作为真实期望值. 值得一提的是,双曲正切函数拟合对于大累计交通量的情况下拟合相关性要显著高于后者,理论上将前者期望值作为真实期望值更为合理,这时双曲正切函数的误检率将能显著提高,而双指数函数的检出率将明显降低.

表 2 双指数函数拟合黑点鉴别结果

累计交通量	不同情况	双指数函数拟合				双曲正切函数拟合			
		相关系数	期望黑点数	非期望黑点数	总黑点数	相关系数	期望黑点数	非期望黑点数	总黑点数
10^6	期望	0.957 7	7	0	7	0.996 3	7	0	7
	第 1 次	0.999 5	6	33	39	0.994 5	5	8	13
	第 2 次	0.997 2	5	13	18	0.999 7	5	13	18
	第 3 次	0.999 3	5	14	19	0.999 0	5	14	19
	第 4 次	0.995 7	6	15	21	0.997 3	6	15	21
	第 5 次	0.996 5	6	19	25	0.999 8	6	19	25
	第 6 次	0.999 4	4	16	20	0.997 4	4	3	7
	第 7 次	0.999 0	5	15	20	0.999 0	5	15	20
	第 8 次	0.998 5	6	21	27	0.990 9	6	21	27
	第 9 次	0.995 2	4	14	18	0.997 4	4	14	18
5×10^6	第 10 次	0.995 4	3	6	9	0.998 6	3	6	9
	期望	0.957 7	7	0	7	0.996 3	7	0	7
	第 1 次	0.976 5	5	0	5	0.998 7	5	1	6
	第 2 次	0.980 4	6	1	7	0.995 8	6	4	10
	第 3 次	0.977 5	6	1	7	0.993 4	6	1	7
	第 4 次	0.975 1	5	0	5	0.997 3	6	0	6
	第 5 次	0.990 7	6	2	8	0.998 2	6	2	8
	第 6 次	0.986 3	6	2	8	0.997 7	6	3	9
	第 7 次	0.983 8	5	2	7	0.998 0	5	2	7
	第 8 次	0.982 0	6	0	6	0.997 8	7	1	8
5×10^7	第 9 次	0.980 6	6	2	8	0.996 7	7	5	12
	第 10 次	0.985 8	6	1	7	0.996 7	6	4	10
	期望	0.957 7	7	0	7	0.996 3	7	0	7
	第 1 次	0.960 2	6	0	6	0.995 6	7	0	7
	第 2 次	0.958 9	5	0	5	0.997 1	6	0	6
	第 3 次	0.965 4	5	0	5	0.995 7	6	0	6
	第 4 次	0.951 8	6	0	6	0.997 0	6	0	6
	第 5 次	0.962 3	6	0	6	0.997 5	6	0	6
	第 6 次	0.957 9	6	0	6	0.996 7	6	0	6
	第 7 次	0.956 2	6	0	6	0.996 9	6	1	7
5×10^7	第 8 次	0.963 0	6	0	6	0.995 8	7	2	9
	第 9 次	0.957 8	6	0	6	0.997 2	6	0	6
	第 10 次	0.946 5	5	0	5	0.996 5	6	1	7

表 3 2 种拟合函数检出率和误检率对比和分析

黑点信息	累计交通量	均值		标准偏差	
		双指数函数拟合	双曲正切函数拟合	双指数函数拟合	双曲正切函数拟合
相关系数 R^2	10^6	0.997 6	0.997 4	0.001 8	0.002 8
	5×10^6	0.981 9	0.997 0	0.004 9	0.001 5
	5×10^7	0.958 0	0.996 6	0.005 6	0.000 7
检出率	10^6	0.714 3	0.700 0	0.15	0.14
	5×10^6	0.814 3	0.857 1	0.07	0.10
	5×10^7	0.814 3	0.885 7	0.07	0.06
误检率	10^6	0.755 2	0.695 0	0.05	0.11
	5×10^6	0.146 4	0.252 0	0.11	0.14
	5×10^7	0	0.050 8	0	0.08

4 应用分析

以沪杭高速公路的上行方向嘉兴段事故数据为基础,应用双曲正切函数对累计频率曲线进行拟合和事故黑点鉴别分析,为比较不同年限的事故黑点鉴别结果,分别在 2004 年 12 月 21 至 2008 年 6 月 20 的事故数据抽取 0.25、1、3.5 a 的事故数据进行应用分析,曲线拟合情况如图 4 所示。

求出拟合曲线上的最小曲率半径的位置作为事故黑点鉴别的临界值,该临界值对应的黑点的事 故临界次数分别为 6 次、18 次和 76 次。根据事故黑点的临界次数,分别鉴定出的事故黑点位置如表 4 所示。

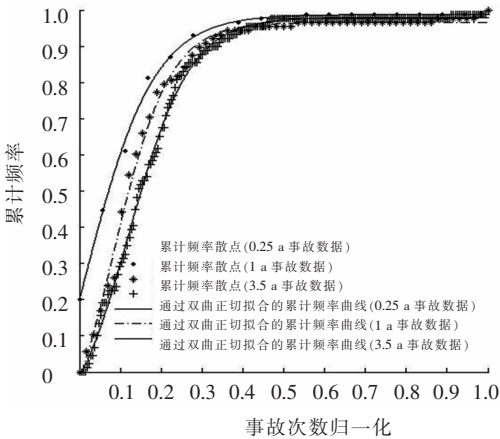


图 4 不同累计交通量作用下累计频率曲线散点图与双曲正切拟合

表 4 不同累计交通量作用下事故黑点鉴别结果

事故统计时间/a	事故黑点
0.25	k75 + 000;k76 + 000;k77 + 000;k99 + 000;k109 + 000;k120 + 000;k115 + 000;k138 + 000
1	k75 + 000;k76 + 000;k99 + 000;k109 + 000;k113 + 000;k117 + 000;k135 + 000;k138 + 000
3.5	k75 + 000;k76 + 000;k99 + 000;k109 + 000;k117 + 000;K135 + 00;k138 + 000

可见,不同累计交通量作用年限下的事故黑点鉴别结果有一定差异性,根据前面的分析可以推断 0.25 a 事故黑点鉴别结果可靠性最低,3.5 a 的事故黑点鉴别结果可靠性最高,1 a 的事故黑点鉴别结果非常接近 3.5 a 的,具有较高的可靠性,其误检率低于 20%,在一定程度上验证了前面的结论.从事故鉴别的结果看,各事故黑点的特征如下:K75 + 000 为大云主线收费站出口;K76 + 000 为大云收费站主线入口;K99 + 000 为沪杭高速与乍嘉苏高速的互通立交;K109 + 000 段为嘉兴服务区;K117 + 000 为桐乡屠甸出口;K135 + 000 为长安服务区,K138 + 000 为长安出口.其中 K99 + 000 处的事故次数达到了 216 次,其次是 K109 + 000 处,事故次数达到 194 次,显著高于其他路

段.可见,对于平原区高速公路,事故黑点位置主要发生在互通立交(特别是枢纽型互通)出入口,服务区出入口和主线收费站出入口等位置。

该方法同样可用于一般等级公路或者其他地区高速公路,对改造前为山区二级公路的 104 国道上虞段 K1545—K1569(1995 年 1 月—1997 年 12 月数据),原太高速公路 K0—K95(2001 年 2 月—2003 年 9 月数据),汉宜高速公路荆州段 K128—K226(1995 年—1999 年数据),太旧高速公路寿阳—旧关段 K446—K508(2001 年 1 月—2003 年 9 月数据)进行黑点鉴别分析,双曲正切函数曲线拟合情况如图 5 所示,黑点鉴别结果如表 5 所示。

从表 5 可知,无论何种情况下,双曲正切拟合

相关系数 R^2 都很高,对于不同的累计交通量具有很好的适应性. 另外统计年限内汉宜高速荆州段上、下行方向的全路段每公里事故数都超过 4 次的最低要求,拟合曲线表现出一定的 S 形,事故黑点鉴别的误检率控制在 20% 左右. 104 国道上海段的全路段每公里事故数接近 4 次,拟合曲线表现出向 S 形转变的趋势,事故黑点鉴别的误检率具有一定的保证. 而原太高速上、下行方向,太旧高速寿阳—旧关段上、下行方向的全路段平均每公里事故次数都明显小于 4 次的最低要求,所以为获得较低的误检率,必须要有更长统计年限的事故数据为基础.

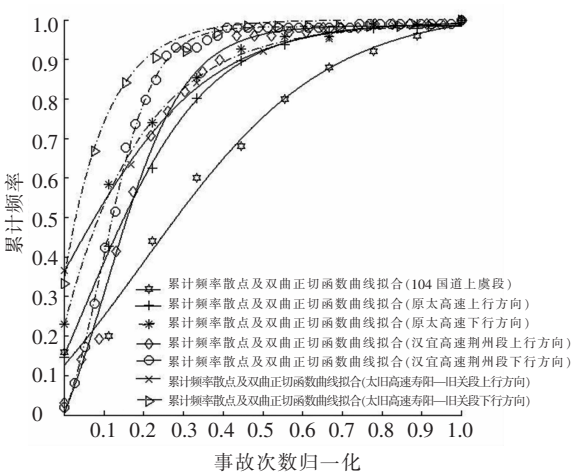


图 5 累计频率曲线散点图与双曲正切拟合

表 5 黑点鉴别结果

路段	黑点个数	全路段平均每公里事故数	曲线拟合相关系数 R^2
104 国道上海段(K1545—K1569)	3	3.36	0.992 0
原太高速上行方向(K0—K95)	10	2.24	0.997 3
原太高速下行方向(K0—K95)	14	1.77	0.998 3
汉宜高速荆州段上行方向(K128—K226)	10	4.89	0.997 2
汉宜高速荆州段下行方向(K128—K226)	7	6.20	0.999 0
太旧高速寿阳—旧关段上行方向(K446—K508)	10	1.30	0.999 0
太旧高速寿阳—旧关段下行方向(K446—K508)	6	1.48	0.997 3

5 结 论

1) 双曲正切函数和双指数函数在累计频率曲线拟合的相关性方面前者要优于后者,并且对于不同的累计交通量具有很好的适应性. 同时,在黑点的检出率方面前者也具备一定的优势,因此双曲正切函数可替代双指数函数进行事故黑点鉴别.

2) 一定事故率下统计年限对事故黑点的误检率有非常重要的影响,当统计年限较低时,事故黑点的误检率非常高,甚至超过 50%. 为将误检率控制在 20% 左右,建议待鉴别道路的平均每公里事故次数不小于 4 次.

参考文献:

[1] ROBERT R V, VEERARAGAVAN A. Hazard rating scores for prioritization of accident prone sections on highways [J]. Transportation Research Record: Journal of the Transportation Research Board, 2004, 1878: 143 - 151.

[2] TARKO A P, KANODIA M. Effective and fair identification of hazardous locations [J]. Transportation Research Record: Journal of the Transportation Research Board, 2004, 1897:64 - 70.

[3] ELVIK R. New approach to accident analysis for hazardous road locations [J]. Transportation Research Record: Journal of the Transportation Research Board, 2003, 1953:50 - 55.

[4] 肖慎,过秀成,宋俊敏. 公路交通事故黑点诊断技术研究[J]. 公路交通科技,2003,20(4):95 - 97.

[5] 裴玉龙,丁建梅. 鉴别道路交通事故多发点的突出因素法[J]. 中国公路学报,2005,18(3):99 - 103.

[6] 孟祥海,盛洪飞,陈天恩. 事故多发点鉴别本质及基于 BP 神经网络的鉴别方法研究[J]. 公路交通科技,2008(03):124 - 129.

[7] 杨轲,方守恩,高国武. 事故多发路段的研究[J]. 上海公路,2000(04):5 - 8.

[8] 方守恩,郭忠印,杨轲. 公路交通事故多发位置鉴别新方法[J]. 交通运输工程学报,2001,1(1):90 - 94.

(编辑 魏希柱)