

文章编号: 1001-0920(2012)07-0991-06

## 基于遗传算法优化和 KD 树的交通流非参数回归预测方法

贾 宁, 马寿峰, 钟石泉

(天津大学 管理与经济学部, 天津 300072)

**摘 要:** 非参数回归预测方法在交通流短时预测中得到了广泛应用. 针对提高搜索速度和关键参数的优化设置两个问题, 提出使用 KD 树作为模式库的存储结构, 能够有效提高搜索速度, 并且能够在实际运行中不断将新发现的交通流模式实时地加入模式库. 提出使用遗传算法对非参数回归中的重要参数进行优化, 实验表明能够得到相对较优的参数设置. 所得研究结果为实时的交通流短时预测系统提供了一种较好的预测方法.

**关键词:** 交通流预测; 非参数回归; 近邻搜索; KD 树; 遗传算法

**中图分类号:** TP273

**文献标识码:** A

## Non-parameter-regression traffic flow forecast method based on KD-tree and genetic optimization

JIA Ning, MA Shou-feng, ZHONG Shi-quan

(College of Management and Economics, Tianjin University, Tianjin 300072, China. Correspondent: JIA Ning, E-mail: jianing.BLGT@gmail.com)

**Abstract:** Non-parametric regression forecasting method is widely used in short-term traffic flow forecasting. To improve its searching speed and parameters optimization, the  $K$ -dimension(KD) tree is used as the structure of the pattern database, which can effectively increase the searching efficiency and allow newly-recognized traffic flow pattern to insert into the pattern database dynamically. Then, the genetic algorithm is used to optimize parameters setting. Experiment results show that better parameters are found by using the algorithm. These results are valuable for the short-term traffic flow forecasting system.

**Key words:** traffic flow forecasting; non-parametric regression; nearest neighbor search;  $K$ -dimension tree; genetic algorithm

### 1 引 言

实时准确的短时交通流预测对于交通管控有着重要的意义. 由于交通流具有很强的不确定性和非线性特性, 传统的基于数学模型的方法难以胜任<sup>[1]</sup>. 因此, 一些不依赖显式数学模型的方法逐渐受到重视, 特别是非参数回归方法由于精度高、鲁棒性好的优点已成为交通流短时预测中最重要的方法之一<sup>[2-5]</sup>.

非参数回归预测实际上是一种基于模式匹配和数据挖掘的方法, 其优点在于它完全是数据驱动, 几乎不需要先验知识, 只需要有足够规模的模式库, 即可做到比较准确的预测. 但基于非参数回归的短时交通流预测方法在实际应用中还存在几个问题:

1) 非参数回归方法的核心是相似模式搜索, 当模式库较大时, 搜索速度成为算法的一个关键性能. 文

献[6]提出了基于动态聚类和散列表的模式库组织方法, 文献[7]提出先将历史数据进行聚类, 然后使用二叉树作为索引的方法, 有效提高了近邻搜索的速度. 但由于交通系统的复杂性, 在实际使用中, 新的交通流模式会不断出现并需要将其加入到模式库中, 如果使用这几种方法, 则必须对数据重新进行聚类处理, 造成额外的开销.

2) 状态分量选择、状态分量的权值及最近邻个数等参数的选取对非参数回归方法的精度有较大的影响. 对于状态分量的选择, 出现了一些效果较好的方法, 如相关系数法、主成分分析和粗糙集等方法<sup>[6,8-9]</sup>. 而对于后两个问题目前研究成果相对不足, 有的文献中提出使用主成分贡献率、相关系数或一些观测统计信息等作为状态分量的权值<sup>[6,8,10-11]</sup>, 但它们

收稿日期: 2011-01-05; 修回日期: 2011-06-16.

基金项目: 国家自然科学基金项目(70971094); 天津市科技支撑计划重点项目(08ZCKFSF01000).

作者简介: 贾宁(1983-), 男, 博士, 讲师, 从事交通流理论、智能交通的研究; 马寿峰(1965-), 男, 教授, 博士生导师, 从事交通系统工程等研究.

只能间接地反映权值对最终预测结果的影响,并不一定是实际最佳的权值.而对于 $K$ 值,目前一般是通过试算的方式进行设置<sup>[5,10,12]</sup>,实用性不佳.

针对这些问题,本文提出了一种基于KD树和遗传算法优化的非参数回归交通流短时预测算法,使用非线性的数据结构——KD树作为模式库的存储结构,相比传统的线性结构有效提高了最近邻搜索的效率,并且新的模式能够较为容易地加入模式库中;同时将遗传算法运用于数据预处理过程,对状态分量的权值和关键参数取值进行优化,能够得到使预测结果较优的参数设置;最后利用实验验证了方法的有效性.

## 2 基于KD树的交通流非参数回归预测方法

### 2.1 交通流非参数回归预测方法

在交通流短时预测中,一般每个预测周期为3~15 min,如果要在第 $t$ 个周期预测下一预测周期某个断面的交通流量,则使用非参数回归方法进行预测的步骤如下:

1) 找出影响该断面下一周期流量的因素 $f_1 \sim f_n$ ,在交通流预测中,这些因素可能包括本断面及上游断面的历史流量、天气、道路情况等.将 $f_1 \sim f_n$ 称作系统的状态分量,这些分量构成了系统状态向量 $(f_1, \dots, f_n)$ .借鉴数据挖掘中的定义,将要预测的流量 $q$ 称作决策属性,于是决策属性的值由系统状态向量决定,即本周期的状态向量 $(f_1, \dots, f_n)$ 决定了待预测的流量 $q$ .

2) 收集交通流数据,如收集到第 $t_i$ 个预测周期的状态向量的值以及 $t_i + 1$ 个周期决策属性的值,则它们可以构成一个交通流模式,形为 $\{f_1(t_i), \dots, f_n(t_i)|q(t_i + 1)\}$ .经过一段时间的运行,将获得的大量交通流模式组织成模式库.

3) 在某个预测周期 $t$ ,首先确定待预测系统状态向量 $F = (f_1(t), \dots, f_n(t))$ ,然后按照一定度量准则,在模式库中搜索状态向量与 $F$ 相似的若干个模式,因决策属性的值是由状态向量决定的,所以可以认为在状态向量相似时,决策属性也将是近似的,于是可根据相似模式中决策属性的值推测出待预测决策属性的值,即 $q(t + 1)$ .

### 2.2 KD树的结构

KD树是由Bentley<sup>[13]</sup>提出的一种用于多维数据快速查找的数据结构,它是二叉排序树的推广,每个节点存储的是一个 $N$ 维数据点.在一棵存储 $N$ 维数据的KD树中,对其中第 $l$ 层的非叶子节点 $T$ ,用 $T_{l_c}$ 和 $T_{r_c}$ 分别表示其左、右孩子节点,且令 $\text{index} = l \bmod N$ ,则有 $T_{l_c}[\text{index}] < T[\text{index}]$ 且 $T_{r_c}[\text{index}] \geq T[\text{index}]$ ,

其中索引运算符 $[i]$ 表示节点所存储 $n$ 维向量的第 $i + 1$ 维,并且 $l$ 和 $i$ 都是从0开始.

构造成一棵KD树的过程是循环地使用数据的每一维划分数据集的过程.根节点依据其第1维将整个数据集划分为2个部分,凡是第1维小于等于根节点第1维的节点为根节点的左子树,其他的作为根节点的右子树;然后在子树中用第2维进行划分,如此进行,直到将所有节点划分完毕.图1展示了1棵2维的KD树的结构,可见,每个节点及其子树对应了状态空间中的一个区域,这一性质对于在KD树上进行最近邻搜索是非常有用的,所以在构建KD树时还需要在每个节点中存储它所对应的空间区域的边界.

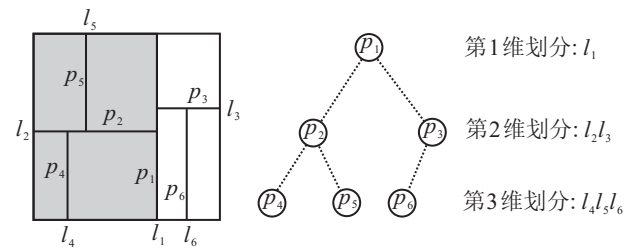


图1 存储2维数据的KD树

与二叉排序树类似,为保证在KD树中进行搜索的效率,在构造树时要保证树的平衡.下面给出了KD树构造算法的伪代码,其中 $T.lchild$ , $T.rchild$ 分别是节点 $T$ 的左、右子树, $d$ 表示递归深度, $S$ 表示数据集,算法返回构造好的KD树根节点:

CreateKDTree( $S, d$ )

if( $S$ 中只有一个节点 node)

return node

else

$i = d \bmod N$

沿着通过 $S$ 内各点第 $i$ 维中值的分割线,

将 $S$ 划分为2个部分,分别记为 $S_1, S_2$ .

生成一个节点 $T$ 用来存储分割集合 $S$ 的中值,并存储它所对应的区域边界.

$T_{l_c} = \text{CreateKDTree}(d + 1, S_1)$ ,

$T_{r_c} = \text{CreateKDTree}(d + 1, S_2)$ .

return  $T$ .

前文已经提到,非参数回归方法的实质是寻找与待预测时刻的系统具有相似状态向量的一组模式,进而利用这些模式中的决策属性的值推算出系统决策属性的预测值.因此在交通流预测中,如果要使用 $\tilde{n}$ 个状态分量来预测决策属性,则需要构造一棵 $\tilde{n}$ 维的KD树,并将决策属性值作为额外数据元素存储在每一个结点中,该值只作为预测之用而在KD树的构建过程中不起作用.

## 2.3 基于KD树的交通流短时预测

### 2.3.1 最近邻搜索策略

在非参数回归方法中,常用的最近邻搜索策略包括2种:核近邻策略和最近近邻策略<sup>[14]</sup>.核近邻策略是指确定一个固定的半径,搜索在这个半径内的所有近邻点;而最近近邻策略是指设置一个固定的数量,在整个模式库中搜索这个数量的近邻点.在交通流预测中,单纯使用两种策略都有所不足,分析如下:

1) 单独使用最近近邻策略.考虑一种极端的状态:模式库中只有1个模式,即使当前系统状态与该模式表示的系统状态相差很大,在匹配时也会找出这个模式作为最近邻点,这是不合理的.面对这种情况,算法的合理反应应当是由于数据不足而无法预测,而不是给出一个误差极大的预测结果.

2) 单独使用核近邻策略.当模式库内存储的模式数量很大时,在规定的搜索半径内能够找到大量的模式,则选择其中若干“最接近”的模式进行预测比使用所有的模式进行预测更为合理.

在本文算法中,综合使用了2种策略:设定一个搜索半径 $R$ 和一个数量 $K$ ,在以待匹配状态为圆心, $R$ 为半径的区域(称作查找区域)内搜索最多 $K$ 个最接近的点.如果区域内不足 $K$ 个点,则使用所有的这些点;如果多于 $K$ 个点,则只保留最近的 $K$ 个.

### 2.3.2 KD树上的最近邻搜索算法

根据上一节给出的最近邻策略,在KD树上进行最近邻搜索的方法如下:递归地对KD树进行搜索,如果某个节点的子树所对应的区域与待查找区域相交,则对子树进行搜索,找出处于查找区域内的点存入最近点集合;如果搜索结束后,列表中所有点的数量大于 $K$ ,则只保留其中 $K$ 个最近的点.

将待搜索区域表示为 $A$ (根据上一节的定义, $A$ 可以用待匹配的数据点本身及半径 $R$ 来描述),最大近邻数为 $K$ ,则在一棵以 $T$ 为根节点的KD树上进行的最近邻搜索的算法伪代码描述如下:

NNSearch( $T, A, K$ )

if  $T$  包含在  $A$  中

    将  $T$  存入最近点列表

else if (region( $T_{lc}$ ) 与  $A$  相交)

    NNSearch( $T_{lc}, A, K$ )

else if (region( $T_{rc}$ ) 与  $A$  相交)

    NNSearch( $T_{rc}, A, K$ )

处理最近邻列表,如果列表中的点数量大于 $K$ ,则只保留 $K$ 个与待匹配数据点距离最近的点.

在上面的算法中,region的作用是获取树中节点对应的空间区域.Friedman等<sup>[15-16]</sup>证明,当需要的近

邻个数 $K$ 相对整个数据集的规模很小时,时间复杂度为 $O(\log N + K)$ ,其中 $N$ 是节点个数.在交通流预测中,由于状态空间很大,每次预测所搜索的区域必定相对很小,否则将影响预测精度.因此可以认为,时间复杂度是 $O(\log N + K)$ .

由于KD树对空间进行的是直线分割,如果使用欧式距离,则由中心和半径决定的查找范围是球形的,在上述算法中判断球与立方体的包含和相交比较复杂;如果使用加权的欧式距离,则判断的难度更大,也更花费计算时间.因此在本文算法中,使用了如下形式的距离定义:

对于两个向量 $X = (x_1, x_2, \dots, x_n)$ ,  $Y = (y_1, y_2, \dots, y_n)$ ,各维的权值分别为 $a_1, \dots, a_n$ ,则 $d(X, Y) = \max(a_1|x_1 - y_1|, \dots, a_n|x_n - y_n|)$ .

由这种距离定义所确定的搜索范围是使用欧氏距离搜索范围的外切立方体,使得在KD树中进行区域的相交和包含判断更为简便.

### 2.3.3 预测算法

设搜索到的 $K$ 个最近邻为 $v_1 \sim v_k$ ,它们与作出预测时刻系统状态的距离为 $d_1 \sim d_k$ ,决策属性(待预测的流量值)分别为 $q_1 \sim q_k$ ,则预测流量为<sup>[12]</sup>

$$q = \sum_{i=1}^K \beta_i q_i, \quad (1)$$

其中

$$\beta_i = d_i^{-1} / \sum_{i=1}^K d_i^{-1}.$$

式(1)体现了“与待匹配状态相似度越高,则其决策属性越具有参考价值”的思想.

### 2.3.4 模式库的调整算法

在短时交通流预测中,面临着各种复杂的交通状况,决定了最初建立的模式库必然不会包含所有的模式,因此,需要在系统的实时运行过程中,不断发现模式库中不存在的模式,并将其加入模式库中.但是如果对这些新模式不加区别地加入模式库,又会造成数据冗余,对搜索速度造成影响.因此将新模式加入模式库的准则是:保证数据有一定的稠密度,又不使数据过分冗余.于是设定两个阈值 $R'$ 和 $K'$ ,对实际交通流数据中的某个模式 $P$ ,首先在模式库中搜寻以 $P$ 为中心, $R'$ 为半径的区域,如果区域内存在的已有模式数少于 $K'$ ,则将 $P$ 加入模式库中,否则不做处理.

用 $P$ 表示待插入的节点, $T$ 表示一个已经建成的KD树的根节点,则向Root所代表的树中插入节点 $P$ 的过程可以用伪代码描述如下:

InsertNode( $P, T$ )

    Parent = SearchNode( $P, T, 0$ )

$l = \text{layer}(\text{Parent}) \bmod N$

if ( $P[l] < \text{Parent}[l]$ )

$T_{lc} = P$

else

$T_{rc} = P$

在上面的过程中, layer 的功能是求某个节点所在的层次, 过程 SearchNode 在一棵已经建成的 KD 树中寻找并返回待插入节点应当被插入的位置, 它需要 3 个参数: 已存在的 KD 树根节点  $T$ , 待插入节点  $P$  和递归深度  $d$ . 伪代码描述如下:

SearchNode( $P, T, d$ )

if ( $T == \text{null}$ )

return 节点  $T$  的父节点

else if ( $P[d \bmod N] < T[d \bmod N]$ )

return SearchNode( $P, T_{lc}, d + 1$ )

else

return SearchNode( $P, T_{rc}, d + 1$ )

与二叉排序树类似, 如果频繁地向 KD 树内插入节点, 则有可能导致 KD 树的结构失去平衡, 查询操作效率降低. 要保证搜索效率, 可以引入与平衡二叉树类似的“平衡度”概念, 随时对树的结构进行调整以保证平衡, 但是由于 KD 树的各层节点不能互换, 它的平衡操作相比平衡二叉树要困难很多, 在最差情况下必须进行整个子树的重建; 另一方面, 除了少数极端情况 (例如退化线性结构), 轻微的不平衡并不会对搜索算法带来大的影响, 因此频繁的平衡调整是不经济也是没有必要的. 为了在查找效率和数据维护之间达到平衡, 本文算法采用如下的策略:

1) 调整的标准: 设定一个“不平衡”阈值  $\lambda$ , 只有当某个子树的左右子树深度之差超过  $\lambda$  时, 才进行平衡操作, 以减小进行平衡操作的频率.

2) 调整的时机: 由于短时交通流预测一般每个周期进行一次, 而每次预测花费的时间相对较短, 两次预测之间有较长的空闲时间对模式库进行检查和调整, 当发现不平衡度超过一定水平时, 才对树进行平衡操作.

### 3 基于遗传算法的参数优化

状态分量的权值以及算法中参数取值与最终的预测结果不存在明确的数学关系, 因此判断设置是否合理最根本的标准是实际预测的效果, 但由于状态分量数量较多, 权值和  $K$  值的取值范围也比较大, 如果使用试算的方法, 效率是非常低下的. 而遗传算法特别适合处理维数高, 且目标函数复杂或没有解析表达形式的问题<sup>[17]</sup>, 因此, 可以使用遗传算法来优化非参数回归中的参数设置. 算法主要过程如下:

1) 状态分量筛选, 这一步可以使用已有文献中的主成分分析、相关系数、粗糙集等方法完成, 找出待优化的若干状态分量.

2) 随机为筛选出的各个状态分量设定权值并随机选择  $K$  值, 组成一条染色体. 由于本文的算法使用了核近邻与最近紧邻的综合搜索策略, 搜索半径  $R$  也需要被编码入染色体中, 并用同样的方式产生出若干个初始染色体组成初始群体.

3) 计算个体的适应度函数, 计算方法如下: 首先利用收集的交通流数据集, 提取出其中的交通流模式, 将其中一部分作为模式库; 然后对每一个染色体, 利用其对应的状态分量权值及  $K$  和  $R$  值, 对另一部分交通流模式数据进行离线预测, 将预测结果与实际结果进行对比. 可以计算如下两个指标以衡量预测效果:

① 平均相对误差

$$E_{\text{ARE}} = \frac{\sum_{i=1}^N |F_i - f_i| / F_i}{N},$$

其中:  $F_i$  表示第  $i$  个周期的预测流量,  $f_i$  表示第  $i$  个周期的实际流量. 这个指标反映了长时间预测的精度.

② 误差分布概率<sup>[18]</sup>  $E_{\text{PPE}}$ , 表示预测误差落在一定范围内的概率. 在本文中, 取误差范围在 20% 以下的概率, 它反映了预测的可靠性.

两个指标所反映的算法性能各有偏重, 在实际使用中适应度函数可根据需求任意选取. 如果要两个指标兼顾, 在一般的遗传算法中只能将适应度函数设计为两个指标的加权平均, 即

$$f = \lambda_1 \cdot (1 - E_{\text{ARE}}) + \lambda_2 \cdot E_{\text{PPE}}.$$

在未来研究中可以尝试使用专门针对多目标优化的其他启发式优化算法.

4) 交叉/变异/选择操作: 属于较成熟的过程, 并非本文重点, 因此不再赘述.

5) 结束准则: 适应度函数不再有明显变化, 或是迭代达到一定次数, 可根据情况选取.

以上方法是根据实际预测结果寻找最佳设置, 给出的结果是针对特定离线数据的优化方案, 所以对于不同应用当中收集的数据集都需要单独优化. 由于遗传算法是一种需要多次迭代的启发式算法, 如果模式库较大, 则算法需要较长的时间. 对于这一问题, 首先, 搜索速度是优化过程的速度瓶颈, 通过 KD 树的使用, 有效地提高了近邻搜索的速度; 其次, 因遗传算法优化属于离线的数据预处理过程, 所以对速度要求并不是很高. 在下一节给出的实验中将会看到, 本算法在实际使用中是可以满足需要的.

## 4 算法实例

### 4.1 搜索速度的对比

首先针对线性结构和KD树结构进行最近邻搜索速度的比较实验, 在实验中, 每个数据点包括10个状态分量, 每个都为0~1之间的实数, 邻域半径为0.1, 取为100. 对不同的数据规模进行搜索速度实验. 实验程序使用C#语言编写, 硬件环境为3.2G CPU, 2G内存, 实验结果见图2.

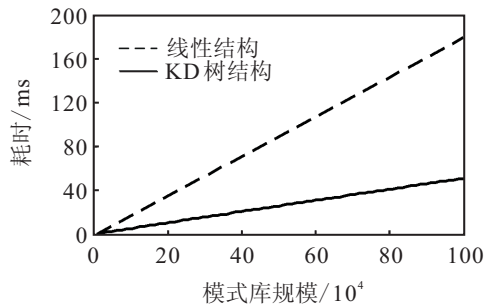


图2 KD树结构与线性结构的搜索速度对比

从图2可以看出, 随着数据规模的增大, 使用KD树进行最近邻搜索的速度提升十分明显.

### 4.2 遗传算法优化效果

本文采用的交通数据来自于明尼苏达德卢斯大学网站 (University of Minnesota Duluth, <http://www.d.umn.edu/tddl/traffic/>), 这些数据采集自明尼苏达市的公路, 实验选取场景如图3所示.

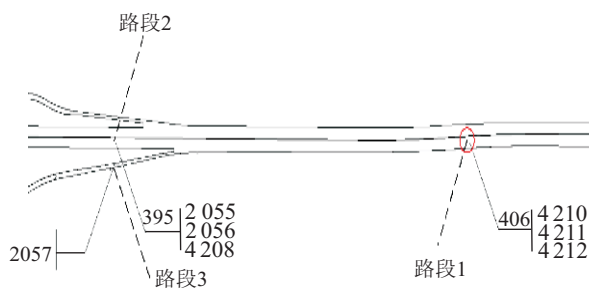


图3 实验道路场景

在所选场景中, 4210, 4211和4212号检测器位于主干公路下游, 2055, 2056和4208号检测器位于主干公路上游, 2057号检测器位于入口匝道. 将3组检测器所在路段分别编号为1, 2, 3, 通过汇总检测数据可以得到每个预测周期经过3个检测点的流量, 并用 $q_1(t)$ ,  $q_2(t)$ ,  $q_3(t)$ 表示它们第 $t$ 个预测周期的流量.

在实验中, 使用2010年1月整月的流量数据作为实验数据集, 将检测点3处的流量作为待预测流量. 预测周期为5 min.

首先使用主成分分析方法, 计算得到预测 $q_3(t)$ 所需的状态分量为 $q_1(t-1)$ ,  $q_2(t-1)$ , 将它们

记为 $\omega_1, \omega_2$ ; 然后以平均相对误差最小为优化目标, 利用第3节给出的优化方法优化 $\omega_1, \omega_2$ 以及 $K, R$ 的值. 优化结果为 $\omega_1 = 1.36, \omega_2 = 0.31, K = 17, R = 13.71$ .

以参数 $K$ 和 $\omega_2$ 为例, 通过对比实验检验优化效果. 其他参数固定采用优化后的结果, 将 $K$ 和 $\omega_2$ 分别取不同的值进行预测, 对比不同参数设置下预测效果的差异. 图4给出了预测结果的变化, 可见, 当参数 $K$ 和 $\omega_2$ 取其他值时, 预测效果变差, 说明优化算法是有效的.

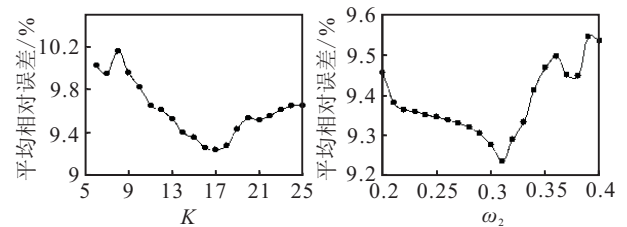


图4  $K$ 和 $\omega_2$ 采用不同值时的预测效果

### 4.3 预测效果

本节的目的是对预测效果进行验证, 评价指标选用上节介绍的平均相对误差和误差分布概率.

在实验中, 使用上一节中相同的场景, 2010年2月的数据为测试数据, 参数设置采用上节得到的优化结果. 为体现预测方法的效果, 令初始模式库为空, 将测试数据逐条输入系统中进行预测, 图5给出了对2月5日, 10日, 15日, 25日数据的预测效果.

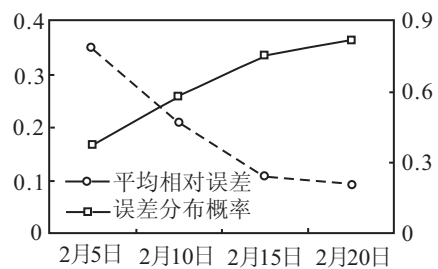


图5 不同日期数据的预测效果

可以发现, 随着时间推移, 预测效果逐渐变好, 4个预测日期的平均相对误差分别为0.346, 0.207, 0.106, 0.091, 误差分布概率分别为0.372, 0.58, 0.752, 0.816. 这是因为预测开始时模式库中预测所需的数据不足, 而随着系统运行, 新的模式不断加入模式库中, 使预测效果得到提高, 体现了模式库实时更新带来的优势.

## 5 结论

本文提出了一种改进的非参数回归交通流短时预测方法, 使用非线性数据结构KD树作为模式库的存储结构, 给出了KD树的构建、基于KD树的最近邻搜索及KD树的调整算法, 实验证明能够有效提高最近邻搜索速度, 并可以较为容易地向模式库中添加数

据. 使用遗传算法对非参数回归中的重要参数, 包括各个状态分量的权值、搜索半径、最近邻个数等进行优化, 并通过实验验证了其优化效果, 对非参数回归交通流预测方法的实际应用具有一定的指导作用.

进一步提高预测方法的精度, 以及如何利用这一方法构建可用于大规模路网实时在线预测的预测系统, 是未来研究的重点.

### 参考文献(References)

- [1] 贺国光, 李宇, 马寿峰. 基于数学模型的短时交通流预测方法探讨[J]. 系统工程理论与实践, 2000, 20(12): 51-56. (He G G, Li Y, Ma S F. Discussion on short-term traffic flow forecasting methods based on mathematical models[J]. Systems Engineering-Theory and Practice, 2000, 20(12): 51-56.)
- [2] Davis G, Nihan N. Nonparametric regression and short-term freeway traffic forecasting[J]. J of Transportation Engineering, 1991, 117(2): 178-188.
- [3] Smith B L, Demetsky M J. Traffic flow forecasting: Comparison of modeling approaches[J]. J of Transportation Engineering, 1997, 123(4): 261-266.
- [4] Smith B L, Williams B M, Keith O R. Comparison of parametric and nonparametric models for traffic flow forecasting[J]. Transportation Research Part C, 2002, 10(4): 303-321.
- [5] Turochy R E. Enhancing short-term traffic forecasting with traffic condition information[J]. J of Transportation Engineering, 2006, 132(6): 469-474.
- [6] 宫晓燕, 汤淑明. 基于非参数回归的短时交通流预测与事件检测综合算法[J]. 中国公路学报, 2003, 16(1): 82-86. (Gong X Y, Tang S M. Integrated traffic flow forecasting and traffic incident detection algorithm based on non-parametric regression[J]. China J of Highway and Transport, 2003, 16(1): 82-86.)
- [7] 张晓利, 贺国光, 陆化普. 基于  $K$ -邻域非参数回归短时交通流预测方法[J]. 系统工程学报, 2009, 24(2): 178-183. (Zhang X L, He G G, Lu H P. Short-term traffic flow forecasting based on  $K$ -nearest neighbors non-parametric regression[J]. J of Systems Engineering, 2009, 24(2): 178-183.)
- [8] 张晓利, 贺国光. 基于主成分分析和组合神经网络的短时交通流预测方法[J]. 系统工程理论与实践, 2007, 27(8): 167-171. (Zhang X L, He G G. The forecasting approach for short-term traffic flow based on principal component analysis and combined NN[J]. Systems Engineering-Theory and Practice, 2007, 27(8): 167-171.)
- [9] 张慧哲, 王坚. 数据挖掘在短时交通流预测模型中的应用研究[J]. 计算机集成制造系统, 2008, 14(4): 690-695. (Zhang H Z, Wang J. Application of data mining on short-term traffic flow forecasting model[J]. Computer Integrated Manufacturing System, 2008, 14(4): 690-695.)
- [10] 张涛, 陈先, 谢美萍, 等. 基于  $K$  近邻非参数回归的短时交通流预测方法[J]. 系统工程理论与实践, 2010, 30(2): 376-384. (Zhang T, Chen X, Xie M P, et al.  $K$ -NN based nonparametric regression method for short-term traffic flow forecasting[J]. Systems Engineering-Theory and Practice, 2010, 30(2): 376-384.)
- [11] Clark S. Traffic prediction using multivariate nonparametric regression[J]. J of Transportation Engineering, 2003, 129(2): 161-168.
- [12] 翁剑成, 荣建, 任福田, 等. 基于非参数回归的快速路行程速度短期预测算法[J]. 公路交通科技, 2007, 24(3): 93-106. (Weng J C, Rong J, Ren F T, et al. Non-parametric regression model based short-term prediction for expressway travel speed[J]. J of Highway and Transportation Research and Development, 2007, 24(3): 93-106.)
- [13] Bentley J L. Multidimensional binary search trees used for associative searching[J]. Communications of the ACM, 1975, 18(9): 509-517.
- [14] Altman N S. An introduction to kernel and nearest-neighbor nonparametric regression[J]. The American Statistician, 1992, 46(3): 175-185.
- [15] Friedman J H, Bentley J L. An algorithm for finding best matches in logarithmic expected time[J]. ACM Trans on Mathematical Software, 1977, 3(3): 209-226.
- [16] Bentley J L, Friedman J H. Data structures for range searching[J]. ACM Computing Surveys, 1979, 11(4): 397-409.
- [17] 李敏强, 寇纪淞, 林丹, 等. 遗传算法的基本理论与应用[M]. 北京: 科学出版社, 2002. (Li M Q, Kou J S, Lin D, et al. Basic theory and application of genetic algorithm[M]. Beijing: Science Press, 2002.)
- [18] 史其信, 郑中为. 道路网短期交通流预测方法比较[J]. 交通运输工程学报, 2004, 4(4): 68-71. (Shi Q X, Zheng Z W. Short-term traffic flow prediction methods comparison of road networks[J]. J of Traffic and Transportation Engineering, 2004, 4(4): 68-71.)