用双切比雪夫方法近似含噪音移动对象轨迹*)

李肖南 丁治明
(中国科学院软件研究所 北京 100080)

摘 要 在移动对象数据库中需要存储大量移动对象的历史轨迹。为了降低存储开销,同时提高轨迹查询的效率,研究者们提出了很多基于时间序列的方法对轨迹序列进行压缩近似及索引。但是这些方法不能用于不精确的轨迹数据。本文针对含噪音的轨迹数据提出了一种新的近似算法。该方法充分利用了轨迹位置数据和速度数据的导数关系,在不增加计算复杂度的情况下,能够更好地处理不精确的轨迹。在相同的压缩比下,用双切比雪夫方法重建的轨迹比现有方法更加接近移动对象的真实轨迹。

关键词 移动对象轨迹,近似,噪音,切比雪夫多项式

Approximating Inaccurate Moving Object Trajectories with Bi-chebyshev Method

LI Xiao-nan DING Zhi-ming
(Institute of Software, Chinese Academy of Sciences, Beijing 100080, China)

Abstract Approximating and compressing inaccurate moving object trajectories is a key topic in moving object database research. This paper proposes a new method, called Bi-Chebyshev, to solve the problem. Different from existing methods, the Bi-Chebyshev method makes use of velocity data as subsidiary information to improve the accuracy of trajectory approximation. The method is based on a quadratic optimization model, to which a numerical solution is provided. In experimental comparison, the Bi-Chebyshev method demonstrates its superiority in both reconstruction accuracy and robustness against noise. The feature coefficients generated by Bi-Chebyshev transformation can be used for indexing purpose in the same way as in other methods. The complexity of computing Bi-Chebyshev coefficients is quadratic on the number of coefficients and linear on the length trajectory data, comparable to other global methods.

Keywords Moving object trajectory, Approximation, Noise, Chebyshev polynomials

1 前言

随着无线通信的广泛应用,移动对象位置的实时跟踪正在成为可能。如何有效地存储和管理大量移动对象的位置信息成为移动对象数据库(Moving Object Database)研究的一个关键问题。一个有效的移动对象数据库不仅要存储移动对象的当前位置,还需要保存历史位置,以便于将来的查询和统计分析。例如,保存某条道路上一个月内各个时段的车辆行驶状况可以为交通流量分析和优化交通管理提供支持。历史轨迹的存储和管理对移动对象数据库研究提出了新的挑战。

一方面,随着时间的推移,移动对象的位置信息不断更新,历史数据增长很快。以交通网络中的车辆管理为例,假设一辆车平均每 20 秒向服务器发送一次位置更新消息,每个位置信息需要占用 20 字节的存储空间。如果将这 20 字节的数据直接存入数据库,那么要在一个同时维护 10000 辆车的移动对象数据库中存储一个季度的历史信息就至少需要 30G 的空间。因此,为了降低存储开销和节省硬件成本,有必要对移动对象的轨迹数据进行压缩和近似。

另一方面,在实际环境中,由于位置采样过程存在误差,服务器得到的位置信息并不精确,甚至存在矛盾。这会给后期的统计分析带来困难。提高采样的密度和精度是提高数据准确性的最有效途径。但是由于采样设备和通信带宽的限制,目前还难以完全依靠硬件技术解决数据不精确的问题,需要通过一些数据处理技术来提高精确性。我们可以将一段轨迹的采样数据看作连续曲线上的离散点,借助连续性特征,对其进行平滑降噪处理。

文献[1]提出了基于动态网络的移动对象数据库模型。该模型通过分离道路的地理信息和移动对象在路网中的抽象位置信息,将二维空间中的移动轨迹转化为一维的时间序列信息,从而使得许多时间序列方面的近似和索引方法可以应用于移动对象的轨迹。

目前国际上对时间序列进行近似索引的主流方法大都遵循一个统一的方式:先从一个较长的原始序列中抽取一组特征数,然后将这一组特征数作为一个向量加入高维索引树中。具体说来有两类方法。

一类是全局方法,如离散傅立叶变换[2,3]、小波变换[4-6]、加权切比雪夫变换[7]。该类方法是将一段长度为 N 的时间序列整体变换到另一个域中(如频域),并选取该域上前 n 个主要分量($n<<N$)作为特征系数。由于每个特征系数均包含来自整个时间序列的信息,因而此类方法称为全局方法。这类方法具有可靠的数学基础,并且实际效果很好。

另一类方法的主要思想是将一个完整的时间序列分割成若干段落,每个段落用一个特征数值近似表示。由于每个特征数值仅与一个局部段落相关,因此该类方法称为局部方法。

*)本文受国家自然科学基金项目资助(项目编号:60573164)。**李肖南** 硕士研究生,研究方向为移动对象数据库;**丁治明** 研究员,主要研究方向为数据库、移动计算、信息检索。

典型的局部方法有固定段长的 PAA[8]法（Piecewise Aggregate Approximation）和自适应调整段长的 APCA[9]法（Adaptive Piecewise Constant Approximation）。局部法的优点在于处理速度非常快，适合联机应用和增量处理。在欧氏距离度量下，局部法的近似结果甚至优于一些复杂的全局方法（如傅立叶变换，加权切比雪夫变换）。

然而，尽管局部方法在处理一般时间序列时有诸多优点，却不适用于移动对象的轨迹。在局部方法中，时间序列每个小段都用一个常量近似表示，如此，移动对象的运动会被解释为在一段时间内静止，然后突然跳转到下一个位置并静止，然后再跳转。这种表示违背了移动轨迹的连续性，给后续的统计分析带来困惑。因此，本文的研究采用了全局方法的框架。现有的全局变换方法都没有考虑噪音对近似效果的影响，并且都是针对单个序列提出的，因而不能直接用于解决本文提出的问题。

在实际应用中，服务器通常不仅获得移动对象的位置信息，还可以得到速度信息。利用文献[1]的原理，速度信息同样可以转换为一维时间序列。从理论上说，速度序列应该是位置序列的一阶导。当这两个序列都存在噪音时，这种导数约束关系对二者的噪音具有中和作用。本文基于该出发点提出了一种新的全局方法——双切比雪夫变换法。该方法是对加权切比雪夫变换法[7]（目前最好的全局方法之一）的改进。与其不同的是，双切比雪夫模型充分利用了第一类和第二类切比雪夫多项式之间的导数关系，将其分别映射为轨迹的位置序列和速度序列，实现了在压缩近似的同时降低噪音。

用双切比雪夫变换得到的特征系数，可以和其它全局方法一样作为一个向量加入高维索引树。因此本文将着重介绍双切比雪夫方法在压缩和近似轨迹数据时的抗噪特性。有关索引的问题，读者可以参考相关文献[2-4,7]。

本文的后续部分组织如下：第 2 节简单回顾了切比雪夫多项式的相关知识，并提出双切比雪夫变换模型。第 3 节给出了求解双切比雪夫特征系数的数值方法。在第 4 节中，我们通过具体的实验和比较来证实双切比雪夫方法在处理含噪音轨迹数据时的优越性。最后是对本文的总结和工作展望。

2 基于双切比雪夫方法的移动对象轨迹近似

如前所述，移动对象数据库通常不仅保存了轨迹的位置信息，也保存了速度信息。如果用一个 K 次多项式去近似表示移动对象的位置曲线，那么该多项式的一阶导则恰好是对速度曲线的近似。反过来，如果用一个 $K-1$ 次多项式去近似速度曲线，那么该多项式的积分恰好是对位置曲线的近似。由此可见，在移动对象轨迹的近似问题上，对位置曲线近似和对速度曲线近似是等价的。

现有方法通常假设位置数据是准确的，因而忽略速度信息，只针对位置信息进行压缩近似处理。在位置数据不准确而速度数据准确的情况下，现有方法也可以用于近似速度曲线，并对近似结果作积分得到位置曲线。但是在实际应用中，位置和速度采样数据往往都不准确，在这种情况下使用任何单一的信息作近似处理都会明显地受到噪音影响，使得近似轨迹与真实的轨迹相差较大。

双切比雪夫模型突破了现有方法只能处理单一数据的限制，综合考虑了位置信息和速度信息，从而在两者都含有噪音的情况下，能够比现有方法更准确地重建移动对象的轨迹。该方法同时使用了第一类和第二类切比雪夫多项式，因此称为双切比雪夫方法。

下面我们首先简单回顾一下切比雪夫多项式的定义及性质，并引入一些相关概念。有关切比雪夫多项式的更多介绍请参见文献[10,11]。

定义 1 第一类切比雪夫多项式的定义为

$$P_k(t)=\cos(k\cdot\arccos t),\ -1\leqslant t\leqslant 1。$$

其中 k 为多项式的次数。对 P_k 求导，我们得到

$$Q_k(t)=P'_k(t)=k\frac{\sin(k\cdot\arccos t)}{\sin(\arccos t)}$$

该式恰好是 $k-1$ 次的第二类切比雪夫多项式乘上常系数 k。

细心的读者可能已经发现 $\{P_k\}$ 和 $\{Q_k\}$ 之间的导数关系与移动对象的位置曲线和速度曲线之间的导数关系恰好形成对应。事实上，双切比雪夫方法正是基于这一对应关系提出的。

在求解双切比雪夫模型的过程中需要用到切比雪夫多项式的一个重要性质——正交性。

定理 1 切比雪夫多项式的正交性

(1) $\{P_k\}$ 关于权函数 $\rho(t)=(\sin(\arccos t))^{-1}$ 正交：

$$\langle P_i,P_j\rangle_\rho=\int_{-1}^{1}\rho\cdot P_iP_j\,dt=\begin{cases}0 & (i\neq j)\\ \pi & (i=j=0)\\ \pi/2 & (i=j>0)\end{cases}$$

(2) $\{Q_k\}$ 关于权函数 $\omega(t)=\sin(\arccos t)$ 正交：

$$\langle Q_i,Q_j\rangle_\omega=\int_{-1}^{1}\omega\cdot Q_iQ_j\,dt=\begin{cases}0 & (i\neq j,i>0,j>0)\\ \dfrac{i^2}{2}\pi & (i=j>0)\end{cases}$$

为了后文的方便讨论，我们在此提出两个有关展开式的概念。

定义 2 设 f 是区间$[-1,1]$上的 K 次多项式，则 f 的 $\{P_k\}$ 展开式定义为 $f=\sum_{k=0}^{K}a_kP_k$。利用定理 1 不难证明，$a_0=\frac{1}{\pi}\langle f,P_0\rangle_\rho$，$a_k=\frac{2}{\pi}\langle f,P_k\rangle_\rho$，$k>0$。类似地有 $\{Q_k\}$ 展开式 $f=\sum_{k=1}^{K}b_kQ_k$，其中 $b_k=\frac{2}{\pi}\langle f,Q_k\rangle_\omega$。

对于区间$[-1,1]$上的一般连续函数 g，也可以进行 $\{P_k\}$ 展开和 $\{Q_k\}$ 展开，展开项往往有无穷多个，但通常只取有限个（前 K 个）展开项即可足够精确地表示 g。

基于以上概念，我们现在介绍用于近似含噪音移动对象轨迹的双切比雪夫模型。如前所述，按照文献[1]提出的方法，移动对象的位置信息和速度信息均可以转化为一维时间序列。这两个一维序列又可以通过相似变换映射到区间$[-1,1]$上。因此为了方便讨论，后文假设所有时间序列及函数均定义在$[-1,1]$区间上。

定义 3 Γ 是$[-1,1]$上的连续可导函数，L 是一个 K 次多项式。定义 Γ 和 L 的双切比雪夫距离 D 为

$$D^2=\int_{-1}^{1}\rho(t)(L-\Gamma)^2dt+\int_{-1}^{1}\omega(t)(L'-\Gamma')^2dt$$

双切比雪夫距离由两部分组成。加号的左边是对 Γ 和 L 之间距离的度量，右半部分是对二者导函数之间距离的度量。这两个度量都属于加权欧氏距离，权函数 ρ 和 ω 分别是第一和第二类切比雪夫多项式的权函数。如果 Γ 是移动对象的位置曲线，Γ' 是对应的速度曲线，L 是对 Γ 的一个近似，那么双切比雪夫距离是对该近似的位置误差和速度误差的综合度

量。

定义 4 对于给定的 Γ，使得双切比雪夫距离 D 最小的 L 即为 Γ 的双切比雪夫近似。L 的 $\{P_k\}$ 展开式中的各项系数称为双切比雪夫系数。

由于平方运算不改变非负函数的单调性，D 最小等价于 D^2 最小，因此，我们只需求解 D^2 的最小化问题就可得到原问题的解。由此，双切比雪夫近似模型实质上是一个二次优化模型，存在唯一的全局最优解，并可以利用偏导数为 0 来求 L。设

$$L(t)=\sum_{k=0}^{K}c_kP_k(t)$$

$$V(t)=L'(t)=\sum_{k=0}^{K}c_kQ_k(t)$$

依次令 $\frac{\partial}{\partial c_k}D^2=0, k=0\cdots K$，可以推得双切比雪夫系数的理论求解式：

$$\begin{cases} c_0=\dfrac{\langle\Gamma,P_0\rangle_\rho+\langle\Gamma',Q_0\rangle_\omega}{\pi}=\dfrac{\langle\Gamma,P_0\rangle_\rho}{\pi} \\ c_k=\dfrac{[\Gamma,P_k]_\rho+\langle\Gamma',Q_k\rangle_\omega}{\dfrac{\pi(1+k^2)}{2}}, \quad k>0 \end{cases} \tag{1}$$

上述系数求解公式有两个主要问题待解决。一是要求 Γ 已知。在实际问题中，Γ 通常是未知的，仅有 Γ 的采样序列。二是含有积分运算。数值积分需要大量迭代，计算开销很大，不适合大规模处理。为了解决这两个问题，我们将在下一节中介绍式(1)的数值解法。

3 求双切比雪夫系数的数值方法

本节提出的双切比雪夫系数的数值解法包含两个关键步骤。一是用采样序列构造分段线性函数来替代式(1)中的 Γ 及 Γ'；另一个是用离散内积代替函数的连续内积(即式(1)中的积分)。该数值方法使得计算一个双切比雪夫系数的复杂度仅为线性。尽管数值方法本身会给计算带来误差，但是第 4 节的实验表明，本文提出的数值算法不会影响双切比方法的优越性。

3.1 分段线性替代函数

定义 5 时间序列 $\{(t_i,x_i)\mid i=0\cdots N\}$ 的分段线性函数为

$$X(t)=\frac{x_i-x_{i-1}}{t_i-t_{i-1}}(t-t_{i-1})+x_{i-1}, t\in(t_{i-1},t_i]$$

其中 t_i 为时间点，x_i 为 t_i 时刻的采样值。

$\{(t_i,l_i,v_i)\mid i=0\cdots N\}$ 表示某段连续轨迹的采样序列。其中 t_i 是采样时间，l_i 是采样时的位置，v_i 是采样得到的速度。设 $\{l_i\}$ 的分段线性函数为 $\Gamma_l(t)$，$\{v_i\}$ 的分段线性函数为 $\Gamma'_l(t)$。当轨迹的采样密度足够大(即 N 足够大)时，分段线性函数能够很好地反映移动对象的实际运动情况，因而可以用 Γ_l 和 Γ'_l 来分别替代双切比雪夫距离公式中的 Γ 和 Γ'。于是，原最小化问题转化为

$$\min_L: D^2=\int_{-1}^{1}\rho(t)(L-\Gamma_l)^2\mathrm{d}t+\int_{-1}^{1}\omega(t)(L'-\Gamma'_l)^2\mathrm{d}t$$

上述替代会把采样数据的误差引入到求解公式中。但是，双切比雪夫距离中的位置曲线和速度曲线存在导数关系，在二者同时存在噪音的情况下，这种约束关系使得求得的近似函数 L 不会过于偏离 Γ_l，L' 也不会过于偏离 Γ'_l。简而言之，两条曲线的导约束关系对于二者的噪音具有中和作用。第 4 节中的实验结果证实了这种中和作用的存在。

3.2 数值积分的离散求解

在双切比雪夫系数的计算公式中存在形如 $\langle f,P_k\rangle_\rho$ 和 $\langle f,Q_k\rangle_\omega$ 的积分。为了能快速计算此类积分，我们借助了三角函数的离散正交性[10]。

定理 2(三角函数的离散正交性)

任给一个自然数 M，令 $\theta_m=\frac{m-0.5}{M}\pi, m=1\cdots M$，则对于所有的 $0\leqslant i,j<M$，有

$$\sum_{m=1}^{M}\cos i\theta_m\cdot\cos j\theta_m=\begin{cases}0 & (i\neq j)\\ M & (i=j=0)\\ M/2 & (i=j\neq0)\end{cases}$$

$$\sum_{m=1}^{M}\sin i\theta_m\cdot\sin j\theta_m=\begin{cases}0 & (i\neq j \text{ or } i=j=0)\\ M/2 & (i=j\neq0)\end{cases}$$

设 $f(t)=\sum_{k=0}^{K}f_kP_k(t), g(t)=\sum_{k=1}^{K}g_kQ_k(t)$ 是某 $[-1,1]$ 区间上连续函数的有限展开式，$t_m=\cos\theta_m$，则由定理 2 和切比雪夫多项式的定义不难证明对于大于 K 的任意自然数 M，

$$\begin{cases} f_0=\dfrac{1}{M}\sum_{m=1}^{M}f(t_m) \\ f_k=\dfrac{2}{M}\sum_{m=1}^{M}f(t_m)\cos k\theta_m, & k>0 \\ g_k=\dfrac{2}{kM}\sum_{m=1}^{M}g(t_m)\sin\theta_m\sin k\theta_m, & k>0 \end{cases}$$

为了简化表示，我们引入如下缩写符号

$$[h,\cos k\theta]_M=\sum_{m=1}^{M}h(t_m)\cos k\theta_m$$

$$[h\sin\theta,\sin k\theta]_M=\sum_{m=1}^{M}h(t_m)\sin\theta_m\sin k\theta_m$$

于是有

$$\begin{cases} f_0=\dfrac{1}{M}[f,\cos 0\theta]_M \\ f_k=\dfrac{2}{M}[f,\cos k\theta]_M, & k>0 \\ g_k=\dfrac{2}{kM}[g\sin\theta,\sin k\theta]_M, & k>0 \end{cases} \tag{2}$$

另一方面，由第 2 小节中的定义 2 知

$$\begin{cases} f_0=\dfrac{1}{\pi}[f,P_0]_\rho \\ f_k=\dfrac{2}{\pi}[f,P_k]_\rho, & k>0 \\ g_k=\dfrac{2}{k^2\pi}[g,Q_k]_\omega, & k>0 \end{cases} \tag{3}$$

综合式(2)和式(3)，不难证明如下定理

定理 3(离散求积定理)

f 是一个连续函数。任给一个自然数 M，令 $\theta_m=\frac{m-0.5}{M}\pi$，$t_m=\cos\theta_m$，则下述等式成立

$$\begin{cases} \langle f,P_k\rangle_\rho=\dfrac{\pi}{M}[f,\cos k\theta]_M \\ \langle f,Q_k\rangle_\omega=\dfrac{k\pi}{M}[f\sin\theta,\sin k\theta]_M \end{cases}, 0\leqslant k<M$$

该定理表明，双切比雪夫系数公式中的积分形式可以通过简单的离散内积精确求值，从而避免了数值积分，大大降低了计算开销。

3.3 算法框架

将定理 3 和第 3.1 节的分段线性函数(Γ_l 和 Γ'_l)代入原系数计算公式，即式(1)，便可得到双切比雪夫系数的数值计算公式

$$\begin{cases} \hat{c}_0 = \dfrac{1}{M}[\Gamma_l, \cos 0\theta]_M \\ \hat{c}_k = \dfrac{[\Gamma_l, \cos k\theta]_M + k \cdot [\Gamma'_l \sin\theta, \sin k\theta]_M}{\dfrac{M(1+k^2)}{2}} \end{cases} \quad (4)$$

综合前述讨论，我们将求解双切比雪夫近似函数的算法框架总结如下：

算法 求双切比雪夫近似函数

```
函数：Solve-BiChebyshev
输入：{(ti,li,vi)|i=0…N}  /* 轨迹时间序列。ti 表示采样时间，li
     和 vi 是 ti 时的位置和速度。 */
     K  /* 最高近似次数 */
输出：{ck}   /* 双切比雪夫系数 */
1. 分别构造{li}和{vi}的分段线性函数 Γl(t)和 Γ'l(t)
2. M←2K   /* 使得 M>K */
3. 新建数组 C，长度为 K+1 /* 存放特征系数 */
4. for k=0→K
5.     由公式 4 计算系数 C[k]←ck
6. end for
7. 返回数组 C
```

步骤 1 的复杂度为 $\Theta(N)$。步骤 4－6 中完成一次系数求值的复杂度为 $\Theta(M)=\Theta(K)$。因此，整个算法的计算复杂度为 $\Theta(K^2+N)$，与目前主要的全局方法[2,7]相当。

4 实验及比较

为了说明双切比雪夫近似方法的有效性，特别是对不精确轨迹数据的噪音中和作用，我们进行了相关实验。本文以加权切比雪夫变换[7]作为实验的比较对象。该方法是目前最好的全局方法之一，并且同样基于切比雪夫多项式。由于加权切比雪夫方法并不是针对含噪音轨迹数据提出的，为了进行公平比较，我们为其增加了平滑去噪的预处理程序，使用的去噪方法是离散余弦变换(DCT)。

一种直观的比较方法是：用两种方法分别对输入轨迹数据进行压缩(保证压缩比相同)，然后用各自的特征系数重建轨迹，再分别计算这两种重建的轨迹与原输入轨迹之间的欧氏误差(我们称为可见误差)，误差越小则说明近似效果越好。然而这种方法不适用于本文的问题背景。本文涉及的含有噪音的轨迹数据，一个近似轨迹越接近含噪音的输入轨迹(即可见误差越小)，只能表明该方法在还原轨迹的同时也“更好”地还原了噪音，也就是说抗噪效果差。

对于含噪音的轨迹的近似问题，一个好的压缩方法应该在相同的压缩比中去除更多的噪音。为了判断近似效果的好坏，应当计算重建的近似轨迹与准确轨迹之间的欧氏误差(真实误差)。真实误差越小说明该近似方法抑制噪音的效果越好。但是实际应用中服务器获取的轨迹都是含有噪音的数据，我们无法获知移动对象的准确轨迹用于计算真实误差，因此本文采用了模拟数据进行实验。

双切比雪夫方法在每一个时间点需要采样两个数据(位置和速度)，因此在相同的采样次数下，它使用的信息量是加权切比雪夫以及其它全局方法的两倍。因此，为了保证比较的公平性，我们为加权切比雪夫方法提供的是经过加倍采样的位置数据，从而使得两种方法获得的数据量相同。我们令两种方法生成相同数目的特征系数，以保证相同的压缩比。模拟实验的流程描述如下：

阶段一：模拟数据的生成

1. 定义轨迹曲线 Γ，对其求导获得相应的速度曲线 Γ'。
2. 确定 $2N$ 个采样点(均匀采样)对 Γ 采样，获得准确的位置序列 X。
3. 给 X 添加噪音，得到含噪音位置序列 X_{2N}。
4. 用上述 $2N$ 个采样点中的奇采样点对 X_{2N} 采样，得到位置序列 X_N。显然，X_N 与 X_{2N} 含有相同的噪音。
5. 用上述奇采样点对速度曲线 Γ' 进行采样，并加入噪音得到含噪音的速度序列 V_N。

阶段二：数据压缩

6. 确定近似多项式的最大次数为 K。相应地，特征系数的个数则为 $K+1$。
7. 用 DCT 平滑 X_{2N}，并用加权切比雪夫方法压缩，得到特征系数集 W。
8. 用双切比雪夫方法压缩 $\{X_N, V_N\}$，特征系数集为 B。

阶段三：重建近似轨迹及误差比较

9. 用 W 重建近似轨迹 L_W，求其与 Γ 的欧氏距离 D_W。
10. 用 B 重建近似轨迹 L_B，求其与 Γ 的欧氏距离 D_B。
11. 比较 D_W 与 D_B。

针对不同的噪音情况，我们分别进行了三组实验。在实验一中，位置序列和速度序列被添加了相同强度的高斯噪音。表 1 显示了在各级噪音强度(用标准差 σ 表示)下，用两种方法重建的轨迹与真实轨迹之间的误差比较。其中，在相同的压缩比下(K 相同)，我们为加权切比雪夫的平滑预处理(步骤 7)尝试了不同的截断频率并选取了最小的误差结果参与比较。数据表明，在强度为 0.5 和 1.5 的高斯噪音条件下，双切比雪夫方法能够更准确地重建轨迹。在其它强度的高斯噪音条件下获得的实验结果与之类似。

表 1 重建误差比较：相同强度的高斯噪音

(a) $\sigma=0.5$

K	5	10	15	20	25	30
W-Cheby	0.32	0.24	0.19	0.18	0.18	0.19
Bi-Cheby	0.22	0.19	0.12	0.09	0.09	0.09

(b) $\sigma=1.5$

K	5	10	15	20	25	30
W-Cheby	0.36	0.36	0.35	0.34	0.37	0.38
Bi-Cheby	0.24	0.22	0.16	0.14	0.14	0.14

注：W-Cheby 表示加权切比雪夫方法；Bi-Cheby 表示双切比雪夫方法。

在实验二中，我们给位置序列添加的高斯噪音强度为 $\sigma_L=1.0$，而给速度序列添加的高斯噪音强度为 $\sigma_V=1.5$。由于速度序列仅被双切比雪夫方法使用，因此该实验实际上是给双切比雪夫方法制造了更为苛刻的条件。从实验结果(见表 2)来看，在上述条件下，双切比雪夫方法仍然能够更好地抵抗噪音。

表 2 重建误差比较：给速度序列添加更强的高斯噪音

K	5	10	15	20	25	30
W-Cheby	0.4	0.34	0.32	0.34	0.36	0.4
Bi-Cheby	0.31	0.31	0.27	0.26	0.27	0.27

在实验三中我们给采样序列中加入低频噪音。部分读者可能会提出，对于低频噪音可以像处理高频噪音那样抑制相关频段。这是一种错误的判断。由于在低频段上分布着大量信号，采用频段抑制会丢失相关频段上的信号，反而加剧近似误差。我们选取了若干具有代表性的低频噪音(如低频正弦噪音)进行实验。表 3 给出的是一个正弦噪音的实验结果(振幅为 0.2，频率 3.5)。在表 3 的比较中可以看出，双切比雪夫

(下转第 288 页)

Software Reuse: a Review of Industrial Studies. Empir Software Eng, 2007, 12: 471-516
[3] Leveson N G. Software Safety: Why, What, and How. ACM Computing Surveys, 1986, 18(2): 125-163
[4] Tokuno K, Yamada S. Stochastic Software Safety/reliability Measurement and Its Application. Annals of Software Engineering, 1999(8): 123-145
[5] Yamada S. Software Reliability/Safety Assessment. J. Japan Society for Safety Engineering, 1990, 33(6): 432-441
[6] Musa J D. Software reliability engineering. McGraw-Hill, 1999
[7] Abbott R J. Resourceful Systems for Fault Tolerance, Reliability, and Safety. ACM Computing Surveys, 1990, 22(1)
[8] Fan I, Filos E. Concurrent Engineering Projects Supported by The European Commission's ESPRIT Programme and Future Trends. Concurrent Engineering-Research and Applications, 2001, 9(2): 166-173
[9] Hurst R. SPMMS-Information Structures in Software Management. Software Engineering Journal, 1986, 1(1): 50-57
[10] Barker K, Dale A, Geroghio L. Management of Collaboration in EUREKA Projects: Experiences of UK Participant Technology Analsys & Strategic Management, 1996, 8(4): 467-482
[11] Lyu M R. Handbook of software reliability engineering. McGraw-Hill, 1996
[12] Bodsberg L, Hokstad P. Transparent Reliability Model for Fault-Tolerant Safety System. Reliability Engineering and System Safety, 1997: 25-38
[13] Ramamoorty C V, Bastani F B. Software Reliability-Status and Perspective. IEEE Ttans. on Software Engineering, 1982, SE-8 (4)
[14] Yamada S. Software reliability model—Fundamentals and applications. Nikka Giren, 1994
[15] Gwandu B A L, Creasey D J. Using Formal Methods in Design for Reliability as Applied to An Electronic System That Integrates Software and Hardware to Perform a Function. Microelectron Reliab., 1995, 35(8): 1111-1124
[16] Leveson N G. Software System Safety and Computers. Addison-Wesley Publishing Company, Inc, 1995
[17] Keene SJ Jr. Assuring software safety//Proc. IEEE Annual Reliability Maintenance Symp. Las Vegas, 1992: 274-279
[18] Yamada S. Reliability/safety evaluation of software. Safety Eng, 1994, 33: 432-441
[19] Kang H G, Sung T. An Analysis of Safety-Critical Digital Systems For Risk-Informed Design. Reliability Engineering and System Safety, 2002, 78: 307-314
[20] Cai K Y. System Failure Engineering and Fuzzy Methodology in Introductory Overview. Fuzzy Sets and Systems, 1996, 83: 113-133

(上接第 251 页)

方法对于低频正弦噪音的去除效果最好，而采用了低频抑制的加权切比雪夫方法最差。对其它低频噪音的实验结果与此类似，因此不再赘述。

表 3 重建误差比较：添加正弦低频噪音

K	5	10	15	20	25	30
W-Cheby	0.3	0.21	0.16	0.17	0.22	0.23
W-Cheby2	0.34	0.3	0.26	0.27	0.29	0.29
Bi-Cheby	0.2	0.19	0.1	0.11	0.11	0.11

注：W-Cheby2 表示采用低频抑制的加权切比雪夫方法。

综上所述，对于含有噪音的移动对象轨迹，双切比雪夫方法能够在相同压缩比下更好地抵抗噪音，用双切比雪夫系数重建的轨迹与真实轨迹更加接近。这是因为，双切比雪夫近似模型综合考虑了位置曲线的误差和速度曲线的误差。通过双切比雪夫距离的约束，位置序列和速度序列之间的导数关系对二者的噪音产生中和作用，从而在一定程度上达到抗噪目的。

结束语 本文针对移动对象轨迹的压缩、近似问题提出了双切比雪夫方法。该方法的主要贡献是：

• 突破了现有方法只能处理单一信息的限制，综合位置和速度信息，建立了二次优化模型。

• 利用两类切比雪夫多项式与移动对象轨迹的导数对应关系，推出了模型的理论解。

• 提出了双切比雪夫系数的快速数值算法，使得计算一个系数的复杂度降为线性。

• 通过实验比较证实了双切比雪夫方法能在压缩轨迹数据的同时更有效地抑制噪音，使重建的轨迹在欧氏距离意义下更加接近真实轨迹。

在进一步的工作中，我们考虑将双切雪夫方法应用到不含速度信息的轨迹采样数据中。一种可能的途径是，用位置序列拟合出一条近似的速度序列，然后再针对这两个序列使用双切比雪夫方法。

参考文献

[1] Ding Zhiming, Güting R H. Managing Moving Objects on Dynamic Transportation Networks//Proc. the 16th Int. Conf. Science and Statistical Database Management. Santorini, Greece, 2004: 287-296
[2] Faloutsos C, Ranganathan M, Manolopoulos Y. Fast subsequence matching in time-Series databases//Proc. the Int. Conf. Management of Data. Minneapolis, Minnesota, USA, 1994: 412-429
[3] Rafiei D, Mendelzon A. Similarity-based queries for time series data//Proc. the Int. Conf. Management of Data. Tucson, Arizona, USA, 1997: 13-25
[4] Chan K P, Fu A W C. Efficient time series matching by wavelets //Proc. the 15th Int. Conf. Data Engineering. Sydney, Australia, 1999: 126-133
[5] Chakrabarti K, Garofalakis M, Rastogi R, et al. Approximate query processing using wavelets//Proc. the 26th Int. Conf. Very Large Data Bases Conference. Cairo, Egypt, 2000: 111-122
[6] Guha S, Harb B. Approximation algorithms for wavelet transform coding of data streams//Proc. the 17th Annual ACM-SIAM Symposium on Discrete Algorithm. Miami, Florida, USA, 2003: 698-707
[7] Cai Y H, Ng R. Indexing spatio-temporal trajectories with Chebyshev polynomials//Proc. Int. Conf. Management of Data. Paris, France, 2004: 599-610
[8] Keogh E, Chakrabarti K, Pazzani M, et al. Dimensionality reduction technique for fast similarity search in large time series databases. Journal of Knowledge and Information Systems, 2001, 3 (3): 263-286
[9] Keogh E, Chakrabarti K, Pazzani M, et al. Locally adaptive dimensionality reduction for indexing large time series databases //Proc. Int. Conf. Management of Data. Santa Barbara, California, USA, 2001: 151-162
[10] Press W H, Flannery B P, Teukolsky S A, et al. Numerical Recipes: The Art of Scientific Computing. 2nd Edition. Cambridge University Press, 1994
[11] Mason J C, Handscomb D. Chebyshev Polynomials. Chapman & Hall, 2003