

基于粗糙集的蛋白质结构分类属性筛选*

敖培, 王川, 张纪

(河南师范大学计算机与信息工程学院, 河南新乡 453007)

摘要: 为了对蛋白质结构进行正确分类, 提出了一种基于粗糙集理论的蛋白质结构分类属性筛选方法。通过多结构比对工具 MAMMOTH-mult 获得条件属性值, 针对分辨矩阵中元素特点提出了分辨矩阵简化方法和改进的属性约简方法。实验以 SCOP 1.71 数据库中结构信息完整的 35 个家族数据集为研究对象, 采用本方法得到 %STRCTCORE 和 %LOOSECORE 两个蛋白质分类属性, 并通过两个属性的 d1a0fa1 与 35 个蛋白质家族、46 626 家族与 35 个结构比对结果散点图可以看出, 将这两个分类属性作为蛋白质结构分类标准, 基本上可以对蛋白质结构进行客观正确的分类。

关键词: 粗糙集; 分辨矩阵; 属性约简; 多结构比对; 蛋白质结构分类

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2014)05-1328-04

doi:10.3969/j.issn.1001-3695.2014.05.010

Selection of protein structure classification characteristics based on rough set

AO Pei, WANG Chuan, ZHANG Ji

(College of Computer & Information Engineering, Henan Normal University, Xinxiang Henan 453007, China)

Abstract: Scientific and reasonable classification attribute is critical for proper protein structure classification. This paper presented a method of selecting protein structure classification attributes based on rough set theory. Attribute values were obtained by multiple structure alignment tool MAMMOTH-mult. It proposed methods of simplified discernibility matrix and improved attribute reduction according to characteristic of elements in discernibility matrix. Data set of 35 families with complete information was taken as the research object in SCOP 1.71 database. The proposed method obtained %STRCTCORE and %LOOSECORE protein classification attributes. From scatter diagrams of structure alignment between d1a0fa1 and 35 families, 46 626 family and 35 structures, it shows that the two protein classification attributes are taken as the standard of protein structure classification, protein structure can be classified correctly.

Key words: rough set; discernibility matrix; attribute reduction; multiple structure alignment; protein structure classification

0 引言

鉴于生物信息尤其是关于蛋白质三维结构信息的急剧增长, 寻求一定的方法比较和对不同的蛋白质进行科学合理的分类显得尤为重要。目前, 应用比较广泛的蛋白质结构分类数据库主要包括 SCOP、CATH 和 FSSP。虽然这些数据库已经在各种蛋白质研究中获得了较好的应用, 但数据库目前的版本在蛋白质结构分类属性的确定上均存在不够客观的缺陷。

粗糙集^[1-5]是一种应用于处理不完备、不相容信息的有效数学工具。该理论将知识看做是论域基于某种等价关系的划分, 利用上、下近似的概念刻画划分间的不相容性。其主要特点是无须提供除数据以外的任何先验知识, 对问题的不确定的描述相较于模糊集的描述是客观的, 且该算法得到的知识是以规则的形式描述的, 易被人们理解和利用。

本文以 SCOP 1.71 数据库中结构信息完整的 35 个家族数据集为研究对象, 通过多结构比对工具 MAMMOTH-mult 获得条件属性值, 并采用所提出的基于粗糙集理论的蛋白质结构分类属性筛选方法, 得到 %STRCTCORE 和 %LOOSECORE 两个

较为客观重要的蛋白质分类属性, 为后续蛋白质结构分类工作奠定基础。

1 粗糙集相关理论

1.1 决策表

在粗糙集中, 信息系统是对知识进行表达和处理的基本工具。决策表示一类特殊而重要的信息系统, 它表示当满足某些条件时, 决策应当如何进行。

定义 1 $S = (U, A, V, f)$ 是一个信息系统, 其中: U 表示对象的非空有限集合, 统称为论域; A 表示属性的非空有限集合; $V = \bigcup_{a \in A} V_a$, V_a 是属性 $a \in A$ 的值域; $f: U \times A \rightarrow V$ 是一个信息函数, 它为每个对象的每个属性赋予唯一的信息值, 即对 $\forall u \in U, \forall a \in A$, 有 $f(u, a) \in V_a$ 。

定义 2 在信息系统 $S = (U, A, V, f)$ 中, 如果 $A = C \cup D$, $D = \{d\}$, 且 $C \cap D = \emptyset$, 子集 C 和 D 分别成为条件属性和决策属性, 那么信息系统 S 则称为决策信息系统, 简称为决策表, 记为 $S = (U, A)$ 或 $S = (U, C \cup D)$ 。

收稿日期: 2013-08-12; **修回日期:** 2013-09-22 **基金项目:** 河南省教育厅科学技术研究重点项目基础研究计划资助项目(13A413506); 河南省教育厅科学技术研究重点项目(12A520028); 河南师范大学新引进博士科研启动费支持课题(qd12136); 河南省软科学研究计划项目(112400450305)

作者简介: 敖培(1979-), 女(蒙古族), 辽宁沈阳人, 讲师, 博士研究生, 主要研究方向为智能控制、电网规划与优化(aopei16@sina.cn); 王川(1976-), 男, 河南人, 副教授, 主要研究方向为数据挖掘; 张纪(1990-), 男, 河南人, 硕士研究生, 主要研究方向为农业信息化。

1.2 属性约简

基于粗糙集的知识获取,主要是通过对决策表的属性约简^[6-10]。

定义 3 在决策表 $S = (U, A)$ 中,对 $a \in A$,如果 $IND(A - \{a\}) = IND(A)$,则称属性 a 在 A 中是不必要的;否则称属性 a 在 A 中是必要的。

定义 4 在决策表 $S = (U, A)$ 中,对 $A = C \cup D, C \cap D = \emptyset$, C 中所有必要属性组成的集合称为属性集 A 的核,记做 $core(A)$ 。

定义 5 在决策表 $S = (U, A)$ 中,对 $P \subseteq A$,如果对每个 $a \in P$ 在 P 中都是必要的,则称 P 是独立的;否则称 P 是相关的。

相关属性集中含有冗余属性,而独立属性集去掉任何一个属性,都将破坏决策信息系统的分类能力。

定义 6 在决策表 $S = (U, A)$ 中,对于 $P \subseteq A$,如果 $IND(P) = IND(A)$,且 P 是独立的,则称 P 是 A 的一个约简。

核是所有约简的子集。

1.3 基于分辨矩阵的属性约简

定义 7 在决策表 $S = (U, C \cup D)$ 中,对于 $U = \{u_1, u_2, \dots, u_n\}$ 为论域, $C = \{a_1, a_2, \dots, a_m\}$ 为条件属性, $D = \{d\}$ 为决策属性, $C \cap D = \emptyset$, 决策 P 的分辨矩阵 $M(S) = (m_{ij})_{n \times n}$ 由式(1)定义:

$$m_{ij} = \begin{cases} a \in C, f(u_i, a) \neq f(u_j, a) \\ \text{当 } f(u_i, d) \neq f(u_j, d) \\ \Phi \\ \text{当 } f(u_i, d) = f(u_j, d) \end{cases} \quad i, j = 1, 2, \dots, n \quad (1)$$

显然,分辨矩阵是一个对角矩阵,其对角线元素全为 Φ ,所以在分辨矩阵中一般只考虑其下三角部分。

定义 8 在决策表 $S = (U, C \cup D)$ 中,对于 $U = \{u_1, u_2, \dots, u_n\}$ 为论域, $C = \{a_1, a_2, \dots, a_m\}$ 为条件属性, $D = \{d\}$ 为决策属性, $C \cap D = \emptyset$, 决策 S 的分辨矩阵 $M(S) = (m_{ij})_{n \times n}$ 由式(2)定义。

性质 1 决策表 $S = (U, A)$ 中,对于 $A = C \cup D, C = \{a_1, a_2, \dots, a_m\}$ 为条件属性, $D = \{d\}$ 为决策属性, $C \cap D = \emptyset$, 决策 S 的分辨矩阵 $M(S) = (m_{ij})_{n \times n}$, 则式(2)成立:

$$\begin{aligned} core(A) &= \bigwedge \{m_{ij} \mid m_{ij} = \Phi \text{ 且 } m_{ij} = a_\tau, a_\tau \in C, \\ &1 \leq \tau \leq m, i = 2, \dots, n, j = 1, \dots, i - 1\} \end{aligned} \quad (2)$$

定义 9 在决策表 $S = (U, C \cup D)$ 中, $C = \{a_1, a_2, \dots, a_m\}$ 为条件属性, $D = \{d\}$ 为决策属性, $C \cap D = \emptyset, M(S) = (m_{ij})_{n \times n}$ 为其分辨矩阵,则决策表 S 的分辨函数由式(3)定义:

$$\Delta = f(a_1, a_2, \dots, a_m) = \bigwedge \{ \forall a \mid a \in m_{ij}, \text{ 其中 } m_{ij} \neq \Phi, \\ i = 2, \dots, n, j = 1, \dots, i - 1 \} \quad (3)$$

分辨函数的每个析取分量即对应于决策表的一个约简。

2 基于多结构比对的蛋白质结构分类

蛋白质序列总是折叠成一定的决定其功能和性质的空间结构。由于在进化过程中结构往往比序列更加保守,所以基于结构信息的比对更能反映蛋白质间的远缘进化关系。基于结构信息的比对思想就是把要比对的两个或多个蛋白质结构放在一起,通过彼此间的旋转、移动,最终达到一个最佳的状态。这个最佳的状态就是相互比对的蛋白质结构的相同部分尽量重叠,以达到最优比对结果。基于结构信息的比对也分为双结构比对和多结构比对^[11-15]。

基于上述概念,蛋白质结构比对问题是一种分类问题。分

类就是寻找元素间具有的自反性 $A \leftrightarrow A$ 、对称性 $A \leftrightarrow B \Rightarrow A \leftrightarrow B$ 和传递性 $A \leftrightarrow B, B \leftrightarrow C \Rightarrow A \leftrightarrow C$ 。双结构比对是基于相似(不相似)的度量,对于两条序列 A 和 B ,若 A 和 B 相似,则可以定义它们之间的距离小于某个值 d_0 ,即

$$A \leftrightarrow B \Rightarrow d(A, B) \leq d_0 \quad (4)$$

基于相似度量的双结构比对满足自反性和对称性,但是由 $d(A, B) \leq d_0$ 和 $d(B, C) \leq d_0$ 无法推出 $d(A, C) \leq d_0$,即基于相似度量的双结构比对是无法满足传递性的。

与双结构比对不同的是,多结构比对中距离函数满足超度量空间不等式:

$$d(A, C) \leq \max\{d(A, B), d(B, C)\} \quad (5)$$

在超度量空间中,距离函数对于所有的 d_0 都具有传递性,即对于两条序列,如果已知 $d(A, B) \leq d_0$ 和 $d(B, C) \leq d_0$,则可以推出 $d(A, C) \leq d_0$,所以基于多结构比对的算法能够满足算法的传递性。因此,本文采用 MAMMOTH-mult 多结构比对软件来获得多种蛋白质结构分类属性。

3 基于粗糙集的蛋白质结构分类属性筛选

为了从比对结果中寻找分类标准,本文采用粗糙集理论来评估、筛选多结构比对结果中的结构属性,用知识约简的办法来寻找适合用来构建分类函数的结构属性。

3.1 分辨矩阵的简化

由定义 7 可知,分辨矩阵是一个大小为 $n \times n$ 的下三角矩阵。如果元素在矩阵对角线上,或是在两个对象的决策属性相同的情况下,对应的矩阵元素为 Φ 。利用分辨矩阵求取决策表约简时,决策属性取值种类越少,矩阵中为 Φ 的元素比例就越大,而 Φ 元素对求取属性约简过程起不到任何作用,因此有必要对分辨矩阵进行简化。本文提出了一种分辨矩阵简化的方法,具体步骤如下:

a) 在不影响其获取决策表的约简和核的能力下,去除分辨矩阵中所有为 Φ 的元素。

b) 设决策属性 $D = \{d_1, d_2, \dots, d_s\}$ 和条件属性 $C = \{a_1, a_2, \dots, a_m\}$, 将决策表中决策属性值相同的对象归并到一起,得到对应不同的决策 s 个子表组成。子表的大小对应分别为 $n_1 \times m, n_2 \times m, \dots, n_s \times m$, 其中 $n_1 + n_2 + \dots + n_s = n$ 。

c) 分别对其中任意两个决策不同的子表求取简化矩阵,矩阵的大小为 $n_i \times n_j (i, j \in s, \text{ 且 } i \neq j)$ 。

综上所述,整个决策表共有 C_s^2 个简化分辨矩阵,所有简化分辨矩阵总大小为 $\sum_{i=1}^{s-1} \sum_{j=i+1}^s n_i \times n_j$ 。显然, $\sum_{i=1}^{s-1} \sum_{j=i+1}^s n_i \times n_j < n \times n$, 即简化分辨矩阵的总大小小于原区分矩阵的大小,有效地节约了区分矩阵的存储空间,节省了每个对象与其余各对象比较决策属性值的时间。

3.2 基于分辨矩阵的改进属性约简算法

定义 10 设 m_1 和 m_2 是分辨矩阵的两个元素,如果 $m_1 \supseteq m_2$, 则称元素 m_1 包含 m_2 。

定理 1 如果分辨矩阵的元素 $m_1 \supseteq m_2$, 则把 m_1 置空后所得的分辨矩阵所生成的属性约简不变。

在利用分辨矩阵的定义计算分辨矩阵的元素时会产生许多重复的元素(即存在相等或者包括的关系),而它们在计算分辨函数时不起作用。基于定义 9 和定理 1, 本文对分辨矩阵约简算法进行改进,基本思想是通过使用一个一维数组存放分

辨矩阵的每一个元素,然后对于每个元素逐一判断是否包含另一元素,如果是,则把该元素置空。具体算法如下:

输入:信息表达系统表 $S = (U, A, V, f)$, $U = \{u_1, u_2, \dots, u_n\}$, $A = \{a_1, a_2, \dots, a_m\}$ 。
输出: $S = (U, A, V, f)$ 的分辨矩阵。
a) $n \leftarrow |U|$, $matrix \leftarrow \Phi$, $m \leftarrow |A|$;
b) $s \leftarrow 1$, $d(s) \leftarrow \Phi$;
c) $i \leftarrow 1$;
d) $j \leftarrow i + 1$;
e) $k \leftarrow 1$;
f) if $f(i, a_k) \neq f(j, a_k)$ then $matrix(i, j) \leftarrow matrix(i, j) + a_k$;
g) $k \leftarrow k + 1$, 如果 $k < m$ 转 f), 否则 $d(s) \leftarrow matrix(i, j)$, $s \leftarrow s + 1$ 转 h);
h) $j \leftarrow j + 1$, 如果 $j \leq n$ 转 e), 否则转 i);
i) $i \leftarrow i + 1$, 如果 $i \leq n - 1$ 转 d), 否则转 j);
j) 对数组 $d(s)$ 的每个元素,逐一判断是否包含另一元素,若是则把该元素置空;
k) 输出分辨矩阵 matrix。

3.3 基于粗糙集的蛋白质结构分类属性筛选方法

基于 3.1 和 3.2 节,本文提出蛋白质结构分类属性的筛选方法,具体步骤如下:

- a) 选取蛋白质结构数据集。
- b) 采用多结构比对工具得到各蛋白质序列比对结果,即信息系统的属性值。
- c) 由于经过聚类处理后的同一类信息对象具有很高的相似性,因此可以将聚类后划为一类信息对象来代替,进而简化信息系统。这里采用 K-means 聚类,基于条件属性将数据集分为 K 类,得到 K 个新的信息对象,新的信息对象的条件属性值为每一类数据集中条件属性值的平均值。
- d) 采用等宽度离散化方法离散化信息系统,并根据定义 7 和性质 1 求取分辨矩阵。
- e) 采用 3.1 节的方法简化分辨矩阵,并求取核。
- f) 采用 3.2 节的方法对分辨矩阵进行属性约简,求得分辨函数,进而得到蛋白质结构分类属性。

4 实验验证

为了验证本文方法,这里选择 SCOP 1.71 数据库为研究对象,从中选取结构信息完整的 35 个家族作为目标数据集。从这 35 个家族中各选取一个蛋白质结构,采用 MAMMOTH-mult 多结构比对工具,分别与事先选定的 d1a0fa1 结构、46 626 家族和每个结构所属家族进行比对,得到包含 % STRCTCORE (100% 的序列都有原子的位置且原子间的间隔在 4Å 以内片段的保守区域所占比例)、% LOOSECORE (至少 66% 的序列有原子的位置且原子间的间隔在 3Å 以内片段的保守区域所占比例)、STRCTRMS (STRCT 保守区域内原子间的均方根偏差)、LOOSERMS (LOOSE 保守区域内原子间的均方根偏差)和 Z-SCORE (模板蛋白质与待测蛋白质之间的匹配度)等五个条件属性和一个决策属性(即属于该类置为 1 或不属于该类置为 0)的 105 条比对结果。采用 K-means 聚类方法将 105 条比对结果按照条件属性聚类,得到 8 个新的信息对象,新的信息对象的条件属性值为每一类数据集中条件属性值的平均值,如表 1 所示。

采用等宽度离散化方法,利用公式 $W = (V_{max} - V_{min})/2$ 离散化信息系统中各连续条件属性,其中 V_{max} 和 V_{min} 分别为连续属性最大值和最小值。对于 STRCTRMS 和 LOOSERMS 两个条件属性,当属性值大于 $(V_{max} - W)$ 置 1, 否则置 0。对于其他三

个条件属性,当属性值大于 $(V_{max} - W)$ 置 1, 否则置 0。得到决策表,如表 2 所示。按照 3.1 节方法得到简化的分辨矩阵如表 3 所示。根据性质 1 得到核为 $\{a, b\}$ 。按 3.2 节方法对简化后的分辨矩阵进行约简,如表 4 所示。

表 1 蛋白质多结构比对结果

对象	条件属性					决策属性
	% STRCTCORE	% LOOSECORE	STRCT RMS	LOOSE RMS	Z-SCORE	
1	89.95	110.83	0.29	0.32	15.11	1
2	16.90	81.43	0.82	1.76	-1.22	0
3	0	139.58	0	1.61	1.26	0
4	96.30	103.85	0.42	0.47	3.27	1
5	12.24	139.58	1.54	2.12	-2.49	1
6	0	107.14	0	1.73	2.15	0
7	24.94	115.71	0.84	1.26	-3.12	1
8	16.03	102.77	0.57	0.82	0.16	0

表 2 决策表

U	a	b	c	d	e	f	U	a	b	c	d	e	f
1	1	1	1	1	1	1	5	0	1	0	0	0	1
2	0	0	0	0	0	0	6	0	0	0	0	0	0
3	0	1	0	0	0	0	7	0	1	0	0	0	1
4	1	0	1	1	1	1	8	0	0	1	1	1	0

表 3 决策表

	2	3	6	8
1	abcde	acde	abcde	ab
4	acde	abcde	acde	a
5	b	Φ	b	bcde
7	b	Φ	b	bcde

表 4 基于分辨矩阵的属性约简

	2	3	6	8
1	Φ	Φ	Φ	Φ
4	Φ	Φ	Φ	a
5	b	Φ	Φ	Φ
7	Φ	Φ	Φ	Φ

由表 4 可以得到分辨函数 $\Delta = ab$, 即这个信息系统有一个约简 $\{a, b\}$ 。根据约简的定义可以知道, % STRCTCORE 和 % LOOSECORE 是该系统可以区分蛋白质结构的两个重要属性。绘制关于 % STRCTCORE 和 % LOOSECORE 属性的 d1a0fa1 与 35 个蛋白质家族、46 626 家族和 35 个结构比对结果散点图, 如图 1~4 所示。

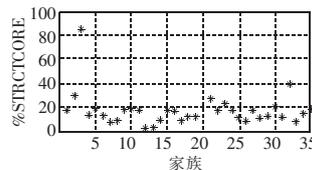


图 1 d1a0fa1 与 35 个家族比对结果中 %STRCTCORE 的散点图

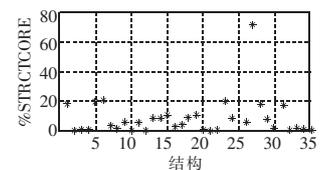


图 2 46 626 家族与 35 个结构比对结果中 %STRCTCORE 的散点图

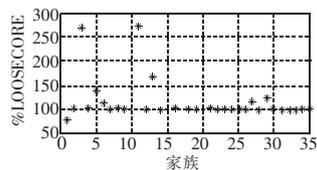


图 3 d1a0fa1 与 35 个家族比对结果中 %LOOSECORE 的散点图

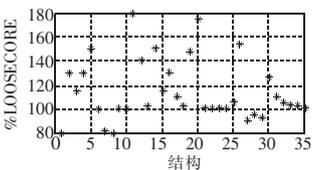


图 4 46 626 家族与 35 个结构比对结果中 %LOOSECORE 的散点图

从图 1 和 2 可以看出,每一组比对中只有一个比对结果中 % STRCTCORE 的值远远大于其他的比对结果,而且这个最大的 % STRCTCORE 正是属于同一家族的蛋白质结构进行比对的结果。因此,印证了 % STRCTCORE 属性区分蛋白质结构的一个重要属性。

从图 3 和 4 可以看出, % LOOSECORE 的信息在图中反映得并不理想, % LOOSECORE 的数值呈现出某些随机性。造成这种结果的原因是在比对过程中采用了一种动态规划的优化过程, % LOOSECORE 是至少有 66% 的序列有原子的位置。如果家族成员比较多时,只需要 66% 的序列原子足够靠近,意味

着增加或减少一条序列结构,对多结构比对的结果影响不大。因此,虽然% LOOSECORE 数值蕴涵着部分的演化信息,是一个非常非常重要的属性,但若将其作为蛋白质结构分类标准还需进行相应的改造。

5 结束语

本文以 SCOP 1.71 数据库中结构信息完整的 35 个家族数据集为研究对象,通过多结构比对工具 MAMMOTH-mult 获得条件属性值,采用粗糙集理论中基于分辨矩阵的属性约简方法来评估、筛选多结构比对结果中的结构属性。经过分析,% STRCTCORE 和 % LOOSECORE 两个分类属性作为蛋白质结构分类标准,基本上可以对蛋白质结构进行正确分类。

参考文献:

- [1] XU Wei-hua, WANG Qiao-rong, ZHANG Xian-tao. Multi-granulation rough sets based on tolerance relations [J]. *Soft Computing*, 2013, 17(7):1241-1252.
- [2] 汤建国,祝峰,余堃,等.粗糙集与其他软计算理论结合情况研究综述[J]. *计算机应用研究*, 2010, 27(7):2404-2410.
- [3] 王学恩,韩崇昭,韩德强,等.粗糙集研究综述[J]. *控制工程*, 2013, 20(1):1-8.
- [4] 王国胤,姚一豫,于洪.粗糙集理论与应用研究综述[J]. *计算机学报*, 2009, 32(7):1229-1246.
- [5] GUO Zhi-lin. Uncertainty analysis for one direction singular rough sets system[C]//Advanced Materials Research. 2013:3170-3173.
- [6] SADAT J N, SALWANI A. Nonlinear great deluge algorithm for rough set attribute reduction[J]. *Journal of Information Science and Engineering*, 2013, 29(1):49-62.
- [7] LI Guo-xiang, XIA Guo-en, PENG Xin. The attribute reduction of regional logistics demand based on rough set theory[J]. *Journal of*

Information and Computational Science, 2013, 10(7):2125-2131.

- [8] WANG Shi-qing, ZHANG Deng-fu, BI Du-yan, et al. Attribute reduction method based on fuzzy rough sets and artificial bee colony algorithm [J]. *Journal of Central South University: Science and Technology*, 2013, 44(1):172-178.
- [9] CHEN Jing-jie, ZOU Ying-huan. A significant impact parameter extraction method based on rough set attribution for reduction fuel consumption estimation[C]//Advanced Materials Research. 2013:1289-1293.
- [10] 杨传健,葛浩,汪志圣.基于粗糙集的属性约简方法研究综述[J]. *计算机应用研究*, 2012, 29(1):16-20.
- [11] AKIHIRO F, MASAHIRO F, YOSHIKAZU M, et al. Consensus substrate sequence for protein-tyrosine phosphatase receptor type Z [J]. *Journal of Biological Chemistry*, 2011, 286(43):37137-37146.
- [12] MINAMI S, SAWADA K, CHIKENJI G. MICAN: a protein structure alignment algorithm that can handle multiple-chains, inverse alignments, C α only models, alternative alignments, and non-sequential alignments[J]. *BMC Bioinformatics*, 2013, 14:24.
- [13] ANDERSON W J, Van DORN L O, INGRAM W M, et al. Evolutionary bridges to new protein folds; design of C-terminal Cro protein chameleon sequences [J]. *Protein Engineering, Design and Selection*, 2011, 24(9):765-771.
- [14] YANG Bing-ru, WU Qu, YING Zhou, et al. Predicting protein secondary structure using a mixed-modal SVM method in a compound pyramid model [J]. *Knowledge-Based Systems*, 2011, 24(2):304-313.
- [15] DANIELS N M, NADIMPALLI S, COWEN L J. Format:correcting protein multiple structural alignments by incorporating sequence alignment[J]. *BMC Bioinformatics*, 2012, 13(1):259.

(上接第 1315 页)CKF 相比 EKF 这一经典非线性滤波方法更适用于系统模型、测量方程等不准确时的滤波,对滤波发散有很好的抑制作用;基于自适应 CKF 的多传感器系统组合滤波方法能够有效地对传感器信息进行融合,提高信息精度;此外,自适应 CKF 的滤波过程相比 EKF 更简单、计算效率更高,更易于工程实现和实际应用。

4 结束语

本文通过加入一组噪声统计估计器和修正函数对容积卡尔曼滤波算法进行了改进,提出了自适应容积卡尔曼滤波算法;针对多传感器系统的组合滤波问题,研究了基于自适应容积卡尔曼滤波的多传感器系统组合滤波结构,并给出了基于容积卡尔曼滤波的多传感器数据融合计算方法;仿真实验结果表明,本文提出的方法对于抑制滤波发散和提高滤波估计的精度,以及多传感器系统的组合滤波等方面都具有良好的效果。本文提出的自适应容积卡尔曼滤波算法假定了噪声的统计特征遵从正态分布这一较为理想的条件,因此在其他条件更为恶劣的情况下,该方法可能不适用;另外,对于多传感器系统中各个传感器相关条件下的组合滤波问题本文也尚未涉及,在未来的工作中将对此进行深入研究。

参考文献:

- [1] 付梦印,邓志红,张继伟. Kalman 滤波理论及其在导航系统中的

应用[M]. 北京:科学出版社,2003.

- [2] 王小旭,潘泉,黄鹤,等.非线性系统确定采样型滤波算法综述[J]. *控制与决策*, 2012, 27(6):801-812.
- [3] 赵琳,王小旭,丁继成,等.基于极大后验估计和指数加权的自适应 UKF 滤波算法[J]. *自动化学报*, 2010, 36(7):1007-1019.
- [4] 胡迪,董云峰.基于自适应 UKF 的传感器故障诊断算法[J]. *北京航空航天大学学报*, 2011, 37(6):639-643.
- [5] 石勇,韩崇昭.自适应 UKF 在目标跟踪中的应用[J]. *自动化学报*, 2011, 37(6):755-759.
- [6] JULIER S J, UHLMANN J K. A new method for the nonlinear transformation of means and covariances in filters and estimators [J]. *IEEE Trans on Automatic Control*, 2000, 45(3):477-482.
- [7] ARASARATNAM I, HAYKIN S. Cubature Kalman filters [J]. *IEEE Trans on Automatic Control*, 2009, 54(6):1254-1269.
- [8] 孙枫,唐李军. Cubature 粒子滤波[J]. *系统工程与电子技术*, 2011, 33(11):2554-2557.
- [9] 王健文,祝海涛,马宏绪.卡尔曼体系下滤波算法计算框架[J]. *数据采集与处理*, 2009, 24(1):61-66.
- [10] 柴霖.联邦滤波器信息分配因子优选问题[J]. *火力与指挥控制*, 2007, 32(8):28-32.
- [11] 顾启泰,王颂.联邦滤波器信息分配系数的优选算法[J]. *中国惯性技术学报*, 2003, 11(6):1-6.
- [12] 熊凯,邱红专.对“联邦滤波器信息分配系数的优选算法”一文的商榷[J]. *中国惯性技术学报*, 2005, 13(2):49-53.
- [13] 付海峰,刘华伟,张宗麟.联邦滤波器公共信息分配方法[J]. *火力与指挥控制*, 2006, 31(8):56-58.