基于概率转移矩阵的氨基酸连接偏好性研究

张 堃， 唐旭清*

（江南大学 理学院，江苏 无锡 214122）

摘要：在 Markov 模型的基础上，提出了状态空间上合并映射的概念，以及合并过程下转移概率的计算方法。在已有氨基酸分类方法的基础上，结合 Markov 模型的概率转移矩阵，对氨基酸连接的偏好性进行了研究。结果表明：同一家族的蛋白质序列的氨基酸连接具有一定的偏好性，这种偏好性与氨基酸的分类有关，从而进一步说明了分类的科学性，同时这种偏好性对氨基酸序列的预测具有一定的作用。

关键词：氨基酸分类；合并映射；概率转移矩阵；偏好性

中图分类号：Q 71；O 29 **文献标识码**：A **文章编号**：1673－1689(2012)01－0106－06

Research on the Connection Bias of Amino Acids Based on Probability Transition Matrix

ZHANG Kun， *TANG Xu-qing**

(School of Science, Jiangnan University, Wuxi 214122, China)

Abstract: In this manuscript, a novel concept of lumping map and a computing method of the transition probability in lumped process were suggested based on Markov model, to investigate the connection bias of amino acids. The results demonstrated that the connection of amino acids had a particular preference which was related to the classification of amino acids, and further verified the scientific of the classification of amino acids. At the same time, the preference would give some help for the prediction of amino acids sequence.

Key words: classification of amino acids, lumping map, probability transition matrix, bias

蛋白质空间结构的所有信息均隐藏在蛋白质的线性结构里面，确切的说，均隐藏在氨基酸序列里面。因此研究蛋白质序列就成了生物信息学研究领域的一个关键问题。目前已经发现的构成蛋白质分子链上的氨基酸类型有 20 种之多，直接研究蛋白质分子的折叠问题有困难，用分类法研究蛋白质结构，已有多种尝试，三联子串（氨基酸）依据其物理和化学特征，或者是依据氨基酸的空间结构特征来进行的不同的分类方式，目前的研究主要集中在几种简化的模型上。K. A. Dill 等人[1]提出的 HP 模型将氨基酸分为 4 类。石秀凡及朱平等人[2]提出的拟氨基酸编码方法将氨基酸分为 16 类，杜晓林等人[3]应用信息聚类的方法将氨基酸分为 5 类，Soumalee Basu 等人[4]在蛋白质序列的混沌游走表达一文中将 20 种氨基酸分为 12 类研究它们的分布情况。分类的依据和偏重不同，分类结果也不同。而这些分类事实上是一种状态合并的问题，即将具有一定关联的对象合并到一个类中，不同的分

收稿日期：2011－01－03

*通信作者：唐旭清（1963－），男，安徽望江人，工学博士，教授，主要从事智能计算，生态系统建模与仿真及生物信息学研究。Email：txq5139@jiangnan.edu.cn

类对应着不同的粒度划分。在实际问题求解中,粒度划分是动态的,常用的氨基酸分类方法都是静态的。Markov 过程是由其转移概率矩阵和初始概率分布构成的,其中的概率转移矩阵描述了其动态性。马氏链预测法[5]是通过对事物不同状态的初始概率及状态之间的转移概率的研究,预测事物的未来状态,在股票预测[6],外汇收益预测[7],基因预测[8-9]等方面都有广泛的应用。作者针对 Markov 模型,结合氨基酸分类方法,对氨基酸连接的偏好性进行了研究,并以木聚糖酶家族[10]的蛋白质序列为例进行了分析。

1 材料与方法

1.1 数据来源

文中的数据来自 Swiss-prot 和 Genebank 中木聚糖酶家族的 6 条蛋白质序列 O43097,P07528,P14768,P23030,P19127,P35811 进行研究。另外文中的相关性分析是通过统计软件 SPSS 来完成的。

1.2 基于粒度下的 Markov 链

1.2.1 合并映射 设 $\{X_n, n \geqslant 0\}$ 有限状态空间 X 上的齐次 *Markov* 链,其中 $X = \{x_1, x_2, \cdots, x_N\}$,如果将 X 中 N 个状态分类成 M 个状态分类成 $C = \{C(1), C(2), \cdots, C(M)\}, (M < N)$。对于给定的分类,建立了一个从 $X = \{x_1, x_2, \cdots, x_N\}$ 到 $Y = \{y_1, y_2, \cdots, y_M\}$ 的一个映射 $\varphi: \forall x_k \in X, \forall y_l \in Y, \varphi(x_k) = y_1 \Leftrightarrow k \in C(l)$,其中映射 φ 称为合并映射或压缩映射,C 称为 X 的一个划分。同时这一过程 $\{Y_n = \varphi(X_n)\}$ 就成为相对于映射 φ 的合并过程[11]。

事实上,这里所给出的合并映射 φ 所起的作用就是给定了原始状态空间 X 上的一个商空间[12-13] $Y = [X]$,以这个商空间作为状态空间(或观测空间)来研究原始马尔科夫链在这较粗状态空间(即 X 的商空间 $[X]$ 下)的性质。在随机线性动力系统中,若给定输入 — 输出动力系统:$x_n + 1 = Ax_n, y_n = Cx_n$,则在什么样的条件下,$y_n$ 具有线性动力系统的性质,特别是当 x_n 是随机变量 X_n 的概率分布时,这个序列与 Markov 链相应,这里 A 是 Markov 链 $\{X_n, n \geqslant 0\}$ 的转移矩阵,而矩阵 C 就是压缩映射 φ,即 y_n 是随机变量 $\varphi(X_n)$ 的概率分布。

1.2.2 合并过程的概率转移矩阵 若 Markov 链 $\{X_n, n \geqslant 0\}$ 的状态空间为 $I = \{1, 2, \cdots, N\}$,设 φ 为从 I 到集合 $Y = \{y_1, y_2, \cdots, y_M\} (M < N)$ 的合并映射,即,$\forall k \in I, \forall y_l \in Y, \varphi(k) = y_l \Leftrightarrow k \in \varphi^{-1}(y_l)$,此时称 Y 为 I 的商空间,$\{Y_n = \varphi(X_n)\}$ 就成为相对于映射 φ 的合并过程。若 $\{X_n, n \geqslant 0\}$ 的初始概率向量为 $X_0 = (\pi_1, \pi_2, \cdots, \pi_N)$,则

$$p(Y_1 = y_t \mid Y_0 = y_s) = \frac{\sum_{i \in \varphi^{-1}(y_s)} \sum_{j \in \varphi^{-1}(y_t)} p_{ij}\pi_i}{\sum_{i \in \varphi^{-1}(y_s)}} (y_s, y_t \in Y), \quad (1)$$

令 $a_{st} = P(Y_1 = y_t \mid Y_0 = y_s)$,矩阵 $A = (a_{st})_{s,t \in Y}$ 为合并过程 $\{Y_n = \varphi(X_n)\}$ 在状态空间 Y 上的转移矩阵。

1.2.3 应用举例 对于一条由 610 个氨基酸构成的蛋白质序列来说,作如下假设,设 610 个位置为 610 个时刻,{A,F,I,L,M,P,V,W,Y,T,S,Q,N,G,C,H,K,R,D,E} 为由 20 种氨基酸构成的状态空间,该状态空间对应于一个号码集合 $I = \{1, 2, \cdots, 20\}$,令 $X(n)$ 表示氨基酸 n 后面所连接的氨基酸种类,显然 $X(n)$ 是一个随机变量,$\{X(n), n = 1, 2, \cdots, 20\}$ 是一个离散参数的随机过程,并且每个氨基酸后面所连接氨基酸与前面的状态无关,只与蛋白质序列本身有关,氨基酸 i 与氨基酸 j 连接的概率与 i 所在的时刻无关,因此氨基酸之间的连接过程可以看成是一个 Markov 过程。于是 Markov 预测模型可以定义为一个三元组 (X, P, π),其中 X 为 20 种氨基酸构成的状态空间,P 为一阶概率转移矩阵,π 为初始分布。定义 X 到 $Y_1 = \{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}, y_{11}, y_{12}\}$ 上的合并映射 φ_1,X 到 $Y_2 = \{y_1, y_2, y_3, y_4\}$ 上的合并映射 φ_2。

$\varphi_1(x_i) = y_i, x_i \in \{\{C\}, \{H\}, \{N\}, \{P\}, \{Q\}, \{W\}, \{Y,F\}, \{A,G\}, \{S,T\}, \{K,R\}, \{D,E\}, \{I,L,V,M\}\}$

$\varphi_2(x_j) = y_i, x_j \in \{\{A,F,I,L,M,P,V,W\}, \{Y,T,S,Q,N,G,C\}, \{H,K,R\}, \{D,E\}\}$

以木聚糖酶家族的 6 条序列 *O*43097,*P*07528,*P*14768,*P*23030,*P*19127,*P*35811 为例。先分别统计出每条序列中两两相连的氨基酸个数,记为 b_{ij},表示氨基酸 i 与氨基酸 j 相连的个数,则 $p_{ij} = \frac{b_{ij}}{\sum_{j \in I} b_{ij}}$ 表示氨基酸 i 与氨基酸 j 相连的概率,显然 $\sum_{j \in I} b_{ij} =$

1，$P_{20\times20}=(p_{ij})_{20\times20}$ 为该预测模型的一阶概率转移矩阵。然后计算出每条序列上氨基酸 i 所占的比例，记为 π_i，则 $\pi=(\pi_i)_{1\times20}$ 为每条序列的初始概率分布。以序列 O43097 为例，该序列中氨基酸 A 与 A 相连的个数为 2，A 与所有氨基酸相连的个数为 22，则 $p_{11}=p_{AA}=\frac{2}{22}=0.090\,909\,1$，同样可以求出一阶概率转移矩阵中的其他元素。另外，该序列的长度为 225，其中氨基酸 A 的个数为 22，则 A 所占的比例为 $\frac{2}{225}=0.097778$，同样可以算出其他 19 种氨基酸所占的比例，该序列的初始分布为(0.097 778，0.008 889，0.053 333，0.044 444，0.031 111，0.142 222，0.017 778，0.031 111，0.017 778，0.053 333，0.004 444，0.057 778，0.035 556，0.035 556，0.04，0.057 778，0.093 333，0.066 667，0.035 556，0.075 556)。用同样的方法可以得到其他序列的一阶概率转移矩阵和初始分布。

这样我们就得到了 6 个初始概率和 6 个 20×20 的一阶概率转移矩阵，通过合并映射 φ_1，φ_2 和公式(1)可得合并后的 12×12，4×4 转移矩阵。

2 结果与分析

2.1 相关性分析

衡量事物之间或变量之间线性相关程度的强弱，并用适当的统计指标表示出来，这个过程就是相关分析[14]。它是研究变量间密切程度的一种常用统计方法。主要分为线性相关分析，偏相关分析，距离相关分析 3 类，作者主要研究线性相关分析。线性相关分析是研究两个变量间线性关系的程度，相关系数是描述这种线性关系程度和方向的统计量，用 r 来描述，若变量 Y 与 X 间是函数关系，则 $r=1$ 或 $r=-1$；如果变量 Y 与 X 间是统计关系，则 $-1<r<1$，一般地，$|r|>0.95$ 存在显著性相关；$|r|>0.8$ 高度相关；$0.5\leqslant|r|<0.8$ 中度相关；$0.3\leqslant|r|<0.5$ 低度相关；$|r|<0.3$ 关系极弱，认为不相关。

相关系数显著性差异的统计意义：由于抽样误差的存在，样本中两个变量间相关系数不为 0，不能说明总体中这两个变量间的相关系数不是 0，因此必须经过检验。检验的零假设是总体中两个变量间的相关系数为 0。SPSS 的相关分析过程给出了该假设成立的概率，公式如下：$t=\frac{\sqrt{n-2}\cdot r}{\sqrt{1-r^2}}$（$r$ 是相关系数，n 是样本观测量数，$n-2$ 是自由度），当相关系数检验的 t 统计量的显著性概率 $p\leqslant0.05$ 时，说明两个变量间相关性显著，通常在概率值上方用“*”表示；当 $p\leqslant0.01$ 时，说明两个变量间相关性非常显著，通常在概率值上方用“**”表示；当 $p\geqslant0.05$ 时，说明两个变量间没有显著的相关关系，只显示概率值。

在 1.2.3 节中得到了 6 条序列的 12×12 和 4×4 概率转移矩阵，以序列 O43097 为例，令 $\alpha=(\alpha_1{}^T,\cdots,\alpha_2{}^T,\alpha_{12}{}^T)^T$，其中 α_i 表示该序列所对应的 12×12 概率转移矩阵中的第 i 列，则 α 为含 144 个分量的列向量，同样的方法可以得到其他五条序列的所对应的列向量，利用 SPSS 软件对这六个列向量进行了相关性分析，结果如表 2.1 所示。按照同样的步骤可以得到 4×4 概率转移矩阵所对应的列向量，相关性分析如表 2.2 所示。

表 1 合并映射 φ_1 下 6 条序列概率转移矩阵的相关性分析

Tab.1 Correlation analysis of the probability transition matrix of the six sequences under the lumping map φ_1

序列		O43097	P07528	P14768	P23030	P19127	P35811
O43097	Pearson Correlation Sig.（2－tailed） N	1 144	0.186(＊) 0.025 144	0.303(＊＊) 0.000 144	0.303(＊＊) 0.000 144	0.373(＊＊) 0.000 144	0.302(＊＊) 0.000 144
P07528	Pearson Correlation Sig.（2－tailed） N	0.186(＊) 0.025 144	1 144	0.397(＊＊) 0.000 144	0.438(＊＊) 0.000 144	0.327(＊＊) 0.000 144	0.423(＊＊) 0.000 144
P14768	Pearson Correlation Sig.（2－tailed） N	0.303(＊＊) 0.000 144	0.397(＊＊) 0.000 144	1 144	0.780(＊＊) 0.000 144	0.508(＊＊) 0.000 144	0.675(＊＊) 0.000 144

续表 1

序列		O43097	P07528	P14768	P23030	P19127	P35811
P23030	Pearson Correlation Sig.（2－tailed） N	0.303(＊＊) 0.000 144	0.438(＊＊) 0.000 144	0.780(＊＊) 0.000 144	1 144	0.483(＊＊) 0.000 144	0.619(＊＊) 0.000 144
P19127	Pearson Correlation Sig.（2－tailed） N	0.373(＊＊) 0.000 144	0.327(＊＊) 0.000 144	0.508(＊＊) 0.000 144	0.483(＊＊) 0.000 144	1 144	0.485(＊＊) 0.000 144
P35811	Pearson Correlation Sig.（2－tailed） N	0.302(＊＊) 0.000 144	0.423(＊＊) 0.000 144	0.675(＊＊) 0.000 144	0.619(＊＊) 0.000 144	0.485(＊＊) 0.000 144	1 144

＊ 在 $\alpha=0.05$ 水平下是显著相关，＊＊ 在 $\alpha=0.01$ 水平下是显著相关.

表 2　合并映射 φ_2 下 6 条序列概率转移矩阵的相关性分析

Tab. 2　Correlation analysis of the probability transition matrix of the six sequences under the lumping map φ_2

序列		O43097	P07528	P14768	P23030	P19127	P35811
O43097	Pearson Correlation Sig.（2－tailed） N	1 16	0.815(＊＊) 0.000 16	0.879(＊＊) 0.000 16	0.949(＊＊) 0.000 16	0.899(＊＊) 0.000 16	0.783(＊＊) 0.000 16
P07528	Pearson Correlation Sig.（2－tailed） N	0.815(＊＊) 0.000 16	1 16	0.866(＊＊) 0.000 16	0.863(＊＊) 0.000 16	0.809(＊＊) 0.000 16	0.739(＊＊) 0.001 16
P14768	Pearson Correlation Sig.（2－tailed） N	0.879(＊＊) 0.000 16	0.866(＊＊) 0.000 16	1 16	0.965(＊＊) 0.000 16	0.973(＊＊) 0.000 16	0.950(＊＊) 0.000 16
P23030	Pearson Correlation Sig.（2－tailed） N	0.949(＊＊) 0.000 16	0.863(＊＊) 0.000 16	0.965(＊＊) 0.000 16	1 16	0.968(＊＊) 0.000 16	0.921(＊＊) 0.000 16
P19127	Pearson Correlation Sig.（2－tailed） N	0.899(＊＊) 0.000 16	0.809(＊＊) 0.000 16	0.973(＊＊) 0.000 16	0.968(＊＊) 0.000 16	1 16	0.929(＊＊) 0.000 16
P35811	Pearson Correlation Sig.（2－tailed） N	0.783(＊＊) 0.000 16	0.739(＊＊) 0.00 116	0.950(＊＊) 0.000 16	0.921(＊＊) 0.000 16	0.929(＊＊) 0.000 16	1 16

＊＊ 在 $\alpha=0.01$ 水平下是显著相关的.

在上述相关性分析表中，Pearson Correlation 表示的是相关系数 r，Sig. 表示的是显著性概率，N 表示的是向量中分量的个数。

从表 1 中可以看出除序列 O43097 和序列 P07528 之间的显著性概率值介于 0.01 和 0.05 之间外，其他序列间的显著性概率均小于 0.01，从表 2 中可以看出，所有序列之间的显著性概率值都小于 0.01，均高度相关，而且相关系数均大于 0.7，相关性都非常显著。这说明在映射 φ_1，φ_2 下 6 条序列都是高度相关的，因此可以将 6 条序列合并处理，得到了合并序列所对应的 20×20 概率转移矩阵及初始分布。

2.2　合并后的概率转移矩阵

根据 1.1 节得到的数据，包括木聚糖酶家族的 6 条蛋白质序列，按照 1.2.3 节定义的合并映射以及公式(1)，得到了 2.1 节中合并序列在 φ_1，φ_2 下的概率转移矩阵，如表 3 和表 4 所示。矩阵中的元素表示两个氨基酸类之间的连接概率，例如，0.025 641 表示的是氨基酸 C 后面连接的氨基酸为 H 的概率是 0.025 641，0.307 692 表示的是氨基酸 C 后

面连接的氨基酸为 A 或者 G 的概率为 0.307 692。

表 3 合并序列在 φ_1 下的概率转移矩阵

Tab. 3 Probability transition matrix of the lumped sequence under the lumping map φ_3

	C	H	N	P	Q	W	YF	AG	ST	KR	DE	ILVM
C	0.051 282	0.025 641	0.076 923	0.025 641	0	0	0	0.307 692	0.205 128	0.076 923	0.025 641	0.205 128
H	0	0.018 868	0.056 604	0.075 472	0.056 604	0	0.169 811	0.113 208	0.207 547	0.037 736	0.056 604	0.207 547
N	0.014 851	0.019 802	0.074 257	0.029 703	0.034 653	0.039 604	0.079 208	0.237 624	0.103 96	0.074 257	0.108 911	0.183 168
P	0	0.018 868	0.066 038	0.028 302	0.009 434	0	0.084 906	0.198 113	0.216 981	0.075 472	0.122 642	0.179 245
Q	0.014 286	0.028 571	0.078 571	0.064 286	0.028 571	0.028 571	0.085 714	0.2	0.135 714	0.107 143	0.035 714	0.192 857
W	0.024 39	0.048 78	0.158 537	0.036 585	0.097 561	0.012 195	0.085 366	0.219 512	0.097 561	0.012 195	0.048 78	0.158 537
YF	0.016 111	0.008 955	0.068 703	0.032 135	0.072 199	0.016 177	0.104 986	0.132 315	0.153 119	0.120 964	0.137 207	0.137 207
AG	0.010 301	0.027 506	0.086 245	0.030 83	0.061 586	0.034 593	0.060 254	0.172 913	0.190 843	0.087 32	0.073 512	0.164 999
ST	0.009 784	0.003 915	0.046 981	0.029 363	0.019 573	0.029 338	0.074 404	0.205 493	0.266 04	0.039 162	0.070 159	0.205 488
KR	0.012 309	0.020 733	0.049 798	0.033 142	0.095 565	0.020 309	0.066 465	0.170 128	0.141 129	0.103 978	0.074 68	0.211 598
DE	0.013 517	0.023 658	0.064 182	0.040 556	0.030 405	0.057 426	0.118 255	0.192 569	0.070 943	0.070 934	0.118 237	0.199 317
ILVM	0.012 348	0.009 267	0.055 566	0.035 215	0.040 592	0.020 477	0.081 412	0.205 573	0.152 496	0.092 708	0.149 989	0.144 058

表 4 合并序列在 φ_2 下概率转移矩阵

Tab. 4 Probability transition matrix of the lumped sequence under the lumping map φ_2

	HKR	DE	CNYGQST	AFILMPVW
HKR	0.112 378	0.071 42	0.421 929	0.394 273
DE	0.094 592	0.118 237	0.368 255	0.418 916
CNYGQST	0.081 718	0.072 99	0.478 133	0.367 159
AFILMPVW	0.111 008	0.127 661	0.428 863	0.332 468

观察表 3 和表 4,可以发现氨基酸之间的连接具有一定的偏好性,通过比较两种分类方式的转移概率,发现该家族的蛋白质序列均偏好使用氨基酸 A,G,S,T,I,L,V,M 均不偏好使用氨基酸 H,K,R,D,E 等,同时还发现氨基酸之间连接的偏好性与氨基酸的分类有关,极性不带电荷 R 基团氨基酸,即{C,N,Y,G,Q,S,T},后面连接{C,N,Y,G,Q,S,T}的概率接近二分之一,而与带正电荷的氨基酸{H,K,R}及带负电荷的氨基酸{D,E}相连的概率则非常小。

3 结 语

氨基酸之间的连接并非随机的均匀的,而是具有一定偏好性的,作者在 Markov 模型的基础上,结合已有氨基酸的分类方法,提出了一种基于概率转移矩阵的氨基酸连接偏好性的研究方法,并以木聚糖酶家族的蛋白质序列为例进行了系统的阐述。研究表明,氨基酸之间的连接具有一定的偏好性,这种偏好性与氨基酸的分类密切相关,同时与密码子使用的偏好性有关。对于木聚糖酶家族而言,极性不带电荷 R 基团氨基酸,后面连接极性不带电荷 R 基团氨基酸的概率接近二分之一,而与带正电荷的氨基酸及带负电荷的氨基酸相连的概率则非常小。这些氨基酸之间连接的偏好性研究对于蛋白质结构预测具有一定的指导意义,一方面可以进一步说明氨基酸分类方式的科学性,同时对蛋白质氨基酸序列的预测有一定的作用,相对于实验室测序、拼装这样预测节省人力物力财力,这将是下一

步研究的主要内容。

参考文献(References):

[1] Lau K F, Dill K A. A lattice statistical mechanics model of the conformation and sequence spaces of proteins[J]. **Macromolecules**,1989,22:3986.

[2] **朱平,高雷,徐振源. 基于拟氨基酸编码方法下的同义密码子的偏好性仍与结合强度密切相关**[J]. **物理学报**, 2009, 6: 714－719.
ZHU Ping, GAO Lei, XU Zhen-yuan. The usage degree of synonymous codon is close correlated with the strength of combination based on the quasi-amino acid coding[J]. **Acta Physica Sinica**,2009,6:714－719. (in Chinese)

[3] **杜晓林,郝玉兰. 氨基酸数量化分类的研究初探**[J]. **生物数学学报**,1994,9(5):105－107.
DU Xiao-lin, HAO Yu-lan. Preliminary study on the quantified Classification of amino acid[J]. **Journal of Biomathematiccs**,1994,9(5):105－107. (in Chinese)

[4] Soumalee B, Archana P, Chitra D, et al. Chaos game representation of proteins[J]. **Journal of Molecular Graphics and Modelling**,1997,15:279－289.

[5] Huseyin P, Wenliang D, Sahin R, et al. Private predictions on hidden markov models[J]. **Artifical Intelligence Review**, 2010,34(1):153－172.

[6] Md. Rafiul H, Baikunth N, Michael K. A fusion model of HMM, ANN and GA for stock market forecasting[J]. **Expert Systems with Applications**,2007,33:171－180.

[7] Dueker M, Christopher J. Neely. Can Markov switching models predict excess foreign exchange returns?[J]. **Journal of Banking & Finance**,2007,31(2):279－296.

[8] **马宝山,朱义胜. 基于隐马尔科夫模型的基因预测算法**[J]. **大连海事大学学报**,2008,34(4):41－44.
MA Bao-shan, ZHU Yi-sheng. Gene-prediction algorithm based on hidden Markov model[J]. **Journal of Dalian Maritime University**,2008,34(4):41－44. (in Chinese)

[9] **张新生,王梓坤. 生命遗传信息中若干数学问题**[J]. **科学通报**,2000,45(2):113－119.
ZHANG Xin-sheng, WANG Zi-kun. Several methematical problems of genetic information of life[J]. **Chinese Science Bulletin**,2000,45(2):113－119. (in Chinese)

[10] **刘亮伟,杨海玉,胡瑜,等**. F/10 **木聚糖酶研究进展**[J]. **食品与生物技术学报**,2009,6:727－732.
LIU Liang-wei, YANG Hai-yu, HU Yu, et al. A review of F/10 xylanase[J]. **Journal of Food Science and Biotechnology**, 2009,6:727－732. (in Chinese)

[11] Leonid G, James L. Markov property for a function of a markov chain: A linear algebra approach[J]. **Linear Algebra and its Applications**,2005,404:85－117.

[12] **张铃,张钹. 问题求解理论与应用:商空间粒度计算理论及应用**[M]. **北京:清华大学出版社**,2007.

[13] Tang X Q, Zhu P, Cheng J X. The structural clustering and analysis of metric based on granular space[J]. **Pattern Recognition**,2010,43:3768－3786.

[14] **朱建平,殷瑞飞**. spss **在统计分析中的应用**[M]. **北京:清华大学出版社**,2007.