

文章编号: 1004-4353(2019)02-0114-03

有序树的计数及其应用

金应烈, 任俊丽

(南开大学 数学科学学院, 天津 300071)

摘要: 利用分拆-组装算法, 构造了含有 $k+1$ 个内点的有序树与森林之间的双射, 讨论了在某些参数限制条件下的有序树的计数, 并在建立 RNA 二级结构与非标号有序树之间的双射基础上, 给出了满足一定参数条件的 RNA 二级结构计数问题的显示闭公式.

关键词: 有序树; RNA 二级结构; 计数

中图分类号: O157.1

文献标识码: A

Enumeration for ordered trees and its applications

JIN Yinglie, REN Junli

(School of Mathematical Sciences, Nankai University, Tianjin 300071, China)

Abstract: We discuss the enumeration of ordered trees under the parameter restrictions by the split-assembly algorithm. Finally we establish a bijection between RNA secondary structures and ordered trees, which is used to enumerate RNA secondary structures under the parameter restrictions, and give the explicit closed formula.

Keywords: ordered tree; RNA secondary structure; enumeration

0 引言

RNA 二级结构的计数问题是计算分子生物学中的研究课题之一, 也是组合数学的一个新兴研究课题. 文献[1-6]研究了满足一定参数条件的 RNA 二级结构计数, 但其结果大多只满足于不同参数条件下的 RNA 二级结构计数的递推关系式、生成函数或近似估计值, 而对完全显示闭公式研究得很少. 文献[7-8]利用不同的对合算法研究了含有 k 个内点的标号有序树的计数. 本文利用标号有序树与森林的对合, 讨论含有 $k+1$ 个内点和 p 个外层内点且外层内点的度不小于正整数 m 的顶点数为 $n+1-k$ 的非标号有序树的计数, 并给出端环长度不小于正整数 m 的 RNA 二级结构计数的完全显示闭公式. 在本文中作如下规定: 对于有序树 T 中的一个内点 u , 若 u 的所有子结点均为叶子点, 则称 u 为外层内点; 否则, 称 u 为内层内点. 高度为 1 的标号有序树称为基本有序树.

1 非标号有序树的计数

定理 1 设含有 $k+1$ 个内点、顶点数为 $n+1-k$ 的标号有序树 T 组成的集合为 A , 由 $k+1$ 个基本有序树组成的顶点数为 $n+1$ 的森林 F_{k+1} 的集合为 B , 其中 $k+1$ 个基本有序树的根结点标号在 $\{1, 2, \dots, n+1-k\}$ 上, 则 A 与 B 之间存在一个双射.

证明 首先, 以集合 A 中的一个有序树 T 构造集合 B 中的一个森林 F_{k+1} . 为方便起见, 称新增加的

顶点 $v_{n+1-k+i}$ ($i=1,2,\cdots,k$) 为星号点,记作 v_i^* .

1) 在有序树 T 的所有外层内点中,找出标号最小的外层内点 v_i ;

2) 从 T 中移走一个以 v_i 作为根、儿子点作为叶子点的标号有序树,并且将其作为森林 F_{k+1} 中的一个基本有序树 T_i ,同时将 T 中原来的顶点 v_i 重新标号为 v_i^* ;

3) 重复以上过程,直至 T 中除根结点以外的所有内点均成为基本有序树 T_i ($i=1,2,\cdots,k$) 的根结点为止.

由以上可得一个由 $k+1$ 个基本有序树所构成的顶点数为 $n+1$ 的森林 F_{k+1} .

其次,以集合 B 中的一个森林 F_{k+1} 构造 A 中的一个有序树 T .

1) 在森林 F_{k+1} 的所有不含星号叶子点的基本有序树中,找出根结点 v_j 的标号 j 为最小的树 T ,并在 F_{k+1} 的所有含星号叶子点的基本有序树中,找出含有最小星号点 v_i^* 的树 T^* ;

2) 将树 T 中的根结点 v_j 与 T^* 中的 v_i^* 合并为一点,且令该点的标号为 j ;

3) 重复以上过程,直至森林 F_{k+1} 变为一个标号有序树.

由以上可证,集合 A 与 B 之间存在一个双射.

从上述的双射构造过程可知,在有序树 T 中若有 p 个外层内点,则在森林 F_{k+1} 中恰有每个根结点的度不小于正整数 m 的 p 个基本有序树,且其叶子点均为非星号点,而其余内层内点所对应的基本有序树的叶子点中至少含有 1 个星号点.

定理 2 设顶点数为 $n+1-k$ 的非标号有序树含有 $k+1$ 个内点和 p 个外层内点,且外层内点的度不小于正整数 m ,则满足条件的非标号有序树的个数为

$$\begin{cases} 1, & k=0; \\ \frac{1}{k+1} \binom{k+1}{p} \binom{k-1}{p-1} \binom{n-pm}{2k}, & k \geq 1. \end{cases}$$

证明 当 $k=0$ 时,根结点外的所有顶点均为叶子点,此时显然只有一个基本有序树.当 $k \geq 1$ 时,由定理 1 知,所求标号有序树 T 的个数与顶点数为 $n+1$ 的森林 F_{k+1} 的个数相等,其中 F_{k+1} 中恰含有 p 个度不小于正整数 m 且其叶子点均为非星号点的基本有序树.因此,若求满足定理条件的标号有序树 T 的个数,只需求顶点数为 $n+1$ 的森林 F_{k+1} 的个数即可.

首先,对森林 F_{k+1} 中的无星号点的 p 个基本有序树的根结点和其余基本有序树的根结点进行标号,其标号的方法数为 $\binom{n+1-k}{p} \binom{n+1-k-p}{k+1-p}$.

其次,假设外层内点所对应的基本树的叶子点有 q 个,则 $pm \leq q \leq n-2k$,且对 p 个基本有序树的叶子点进行标号的方法数为 $q! \binom{n-2k}{q} \binom{q-pm+p-1}{q-pm}$.

再次,将 k 个星号点进行排列后分给剩余的 $k+1-p$ 个基本有序树的叶子点,然后将 $n-2k-q$ 个非星号点再进行排列后分给 $2k+1-p$ 个位置上,其标号的方法数为

$$k! (n-2k-q)! \binom{k-1}{p-1} \binom{n-p-q}{2k-p}.$$

由 $pm \leq q \leq n-2k$ 可知,标号森林 F_{k+1} 的个数为

$$\sum_{q=pm}^{n-2k} \binom{n+1-k}{p} \binom{n+1-k-p}{k+1-p} q! \binom{n-2k}{q} \binom{q-pm+p-1}{q-pm} k! (n-2k-q)! \binom{k-1}{p-1} \times \binom{n-p-q}{2k-p} = \frac{(n+1-k)!}{k+1} \binom{k-1}{p-1} \binom{k+1}{p} \binom{n-pm}{2k}.$$

由此知所求非标号有序树的个数为 $\frac{1}{k+1}\binom{k-1}{p-1}\binom{k+1}{p}\binom{n-pm}{2k}$.

2 RNA 二级结构的计数

引理 1^[3] 设在集合 $\{1,2,\cdots,n\}$ 上,具有 k 个碱基对且所有端环长度不小于正整数 m 的 RNA 二级结构的个数为 $H_n(k)$, 则有:

$$\begin{cases} H_{n+1}(k) = H_n(k) + \sum_{i=m}^{n-1} \sum_{j=0}^{k-1} H_i(j) H_{n-i-1}(k-1-j), & k \geq 1, n > m; \\ H_n(k) = 0, & k \geq 1, n \leq m; \\ H_n(0) = 1, & n \geq 0. \end{cases}$$

下面讨论端环长度不小于正整数 m 的 RNA 二级结构计数的完全显示闭公式.

定理 3 设在集合 $\{1,2,\cdots,n\}$ 上含 k 个碱基对和 p 个端环的 RNA 二级结构的集合为 A , 含有 $k+1$ 个内点、 $n-2k$ 个叶子点和 p 个外层内点的顶点数为 $n+1-k$ 的有序树的集合为 B , 则在 A 与 B 之间存在一个双射.

证明 首先,以集合 A 中的 1 个 RNA 二级结构 S 构造集合 B 中的 1 个有序树 T . 令顶点 u 为树 T 的根.

- 1) 若 (i,j) 是 S 的一个碱基对,且其包含的结构为一个分支,则把点 i 作为 u 的儿子内点;若 i 为一个外点,则将点 i 作为 u 的叶子点,并按 S 中原来的左右顺序进行排列;
- 2) 若 i 是 u 的儿子内点,则去掉碱基对 (i,j) ,并按步骤 1) 把 (i,j) 内部的所有外点和分支作为顶点 i 的儿子结点;
- 3) 重复步骤 2),直至考虑完所有的碱基对为止.

其次,以集合 B 中的一个有序树 T 构造集合 A 中的一个 RNA 二级结构 S . 令 $T(u)$ 为根结点 u 的所有儿子结点的集合.

- 1) 对 $\forall v \in T(u)$, 如果 v 为 T 中的叶子点,则 v 在 RNA 二级结构中对应的点就是外点;如果 v 为内点,则 v 对应的点就是 RNA 中的一个分支,并按照它们原来的左右顺序插入到 RNA 二级结构中;
- 2) 对 $T(u)$ 中的内点 p ,将 $T(p)$ 中的所有点按照步骤 1) 均插入到 (p,q) 分支的内部;
- 3) 重复步骤 2),直至考虑完所有的顶点为止.

由以上可证,集合 A 与 B 之间存在一个双射.

由定理 3 可得 RNA 二级结构与非标号有序树的对应关系,见表 1. 由表 1 和定理 2 可得定理 4.

表 1 RNA 二级结构与非标号有序树的对应关系

RNA 二级结构	非标号有序树
k 个碱基对	含 $k+1$ 个内点
含 $n-2k$ 个自由基	含 $n-2k$ 个叶子点
含 p 个端环	含 p 个外层内点
端环长度为 m	外层内点的度为 $m+1$

定理 4 在集合 $\{1,2,\cdots,n\}$ 上,具有 k 个碱基对和 p 个端环且所有端环长度不小于正整数 m 的

RNA 二级结构的个数为
$$\begin{cases} 1, & k=0; \\ \frac{1}{k+1}\binom{k+1}{p}\binom{k-1}{p-1}\binom{n-pm}{2k}, & k \geq 1. \end{cases}$$

推论 1 在集合 $\{1,2,\cdots,n\}$ 上,具有 k 个碱基对且所有端环长度不小于正整数 m 的 RNA 二级结构

的个数为
$$\begin{cases} 1, & k=0; \\ \frac{1}{k+1}\sum_{p=1}^k\binom{k+1}{p}\binom{k-1}{p-1}\binom{n-pm}{2k}, & k \geq 1. \end{cases}$$

注 推论 1 的结论就是引理 1 递推关系式的完全显示闭公式.

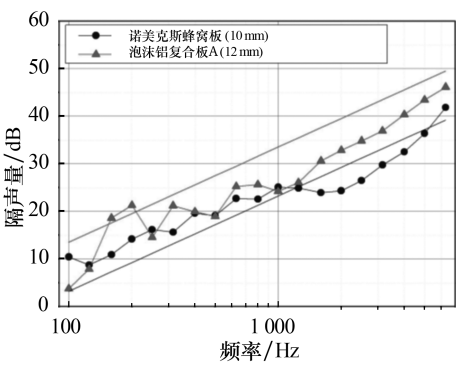


图 7 泡沫铝复合板与诺美克斯蜂窝板的隔声量对比

3 结论

将泡沫铝复合板的隔声量与胶合板和诺美克斯蜂窝板的隔音量进行了对比,结果表明泡沫铝复合板的隔音性能优于这两种材料,且泡沫铝复合板的隔声性能随着面密度和厚度的增加而增强. 本文所得结果可以为泡沫铝复合板的隔声应用提供参考. 本文在实验中没有考虑泡沫铝孔隙的大小对隔声量的影响,今后将进一步研究泡沫铝孔隙大小对隔声性能的影响.

参考文献:

[1] YANG Donghui, YANG Shangrun, WANG Hui, et al. Compressive properties of cellular Mg foams fabricated by melt-foaming method[J]. Materials Science and Engineering, 2010, A527 (21): 5405-5409.

[2] 黄小清,刘逸平,汤立群,等. 基于相对即时密度的泡沫铝材料力学性能研究[J]. 实验力学,2004,19 (2):170.

[3] 王耀琦. 多层泡沫铝填充胀环复合机构缓冲特性研究[D]. 太原:中北大学,2018.

[4] 陈明营,纪箴,贾成厂,等. 泡沫铝及其复合材料的研究进展[J]. 粉末冶金技术,2019,37(1):68-73.

[5] 薛涛. 多孔金属材料泡沫铝的发展[J]. 机械工程材料,1992(1):4-6.

[6] 姜燕坡,白国锋,隋富生,等. 多层复合材料的声学性能研究[J]. 环境工程,2012(S1):211-213.

[7] 声学 建筑和建筑构件隔声测量第 3 部分:建筑构件空气声隔声的实验室测量:GB/T 19889. 3—2005 [S]. 北京:中国标准出版社,2006.

(上接第 116 页)

参考文献:

[1] WATERMAN M S. Combinatorics of RNA hairpins and cloverleaves[J]. Studies in Applied Mathematics, 1978, 60:91-96.

[2] SCHMITT W R, WATERMAN M S. Linear trees and RNA secondary structures[J]. Discrete Applied Mathematics, 1994,51:317-323.

[3] HOFACKER I L, SCHUSTER P, STADLER P F. Combinatorics of RNA secondary structure[J]. Discrete Applied Mathematics, 1998,88:207-237.

[4] LIAO B, WANG T M. General combinatorics of RNA secondary structures[J]. Mathematical Biosciences, 2004, 191:69-81.

[5] WANG W W, ZHANG M, WANG T M. Asymptotic enumeration of RNA secondary structure[J]. J Math Anal Appl, 2008,342:514-523.

[6] CLOTE P. Combinatorics of saturated secondary structures of RNA[J]. Journal of Computational Biology, 2006, 13:1640-1657.

[7] CHEN W Y C. A general bijective algorithm for trees[J]. Proc Natl Acad Sci USA, 1990,87(24):9635-9639.

[8] LIU C L, WANG Z H. A bijection between ordered trees and bicoloured ordered trees[J]. Discrete Mathematics, 2009,309:1417-1421.