

# 1 背景知识

在我们详细介绍算法细节之前，让我们先了解一下关于图分析计算的所需的基本背景。这章包含 (i) 图核函数，(ii) 图小波分析两部分。

## 1.1 图

一个标号图可以被一个有限的节点集合  $V$  和一个有限的边集合  $E \in V \times V$  所描述。在大多数应用中，图都是有标号的。这些标号都是从一个标号集选取的，我们用一个标号函数  $\lambda : V \cup E \rightarrow \Sigma$  来给各个节点和边分配标号。在标号点图中只有点有标号，同样，在标号边图中只有边有标号。在全标号图中，边点都有标号。如果用一种特殊的标号来表示未标号的点和边，那么标号点图和标号边图都可以被看做是全标号图的特殊形式。所以在此文中我们只考虑全标号图来简化问题而又不失一般性。对于标号集  $\Sigma$  我们并不指定具体结构，可以是一个字段，一个向量，也可以是很简单的是个集合。以下我们约定，一幅图用一个四元组  $G = (V, E, \Sigma, \lambda)$  表示， $V, E, \Sigma, \lambda$  都如上文所述。如果一幅图  $G = (V, E, \Sigma, \lambda)$  和另一幅图  $G' = (V', E', \Sigma', \lambda')$  有 1-1 映射的关系  $f : V \rightarrow V'$ ，那么图  $G$  就是  $G'$  的子图，用  $G \in G'$  表示。1-1 映射可以有这么几种

- 对于所有  $v \in V, \lambda(v) = \lambda'(f(v))$
- 对于所有  $(u, v) \in E, (f(u), f(v)) \in E'$
- 对于所有  $(u, v) \in E, \lambda(u, v) = \lambda'(f(u), f(v))$

换言之，如果一幅图保持着另一幅图的所节点标签，边关系，和边标签，那么这副图就是另一幅图的子图。一个通路 (walk) 是一个节点列表  $v_1, v_2, \dots, v_n$ ，对于所有  $i \in [1, n-1], v_i$  和  $v_{i+1}$  有边连接。一个路径 (path) 是一个不含重复节点的通路，即对于所有  $i \neq j, v_i \neq v_j$ 。

## 1.2 再生 Hilbert 空间

对于大量图数据的分析，核函数是一个很强大的处理工具。核函数的优势在于它无需精确计算对应点对就可以把一组数据放入一个高维的 Hilbert 空间。我们用一个叫做核的函数来处理这个过程。一个二元函数  $K : X \times X \rightarrow \mathbb{R}$  如果符合以下公式，则它是一个半正定函数。

$$\sum_{i,j=1}^m c_i c_j K(x_i, x_j) \geq 0 \quad (1)$$

上式中  $m \in \mathbb{N}$ , 例子  $x_i \in X (i = [1, n])$ , 系数集  $c_i \in \mathbb{R} (i = [1, n])$ 。另外, 如果  $x, y \in X, K(x, y) = K(y, x)$  那么这个二元函数就是对称的。一个对称半正定函数保证存在一个 Hilbert 空间  $\mathcal{H}$  和一个映射  $\Phi : X \rightarrow \mathcal{H}$ , 例如

$$k(x, x') = \langle \Phi(x), \Phi(x') \rangle \quad (2)$$

对于所有的  $x, x' \in X$ 。  $\langle x, y \rangle$  表示  $x$  与  $y$  的内积。这个结果就是我们所熟悉的 Mercer 定理。一个对称半正定函数又称为 Mercer 核函数简称为核函数。通过将数据空间变为 Hilbert 空间, 核函数提供了一种对包括图在内的不同数据的统一分析环境。即使一开始的数据空间根本不像一个向量空间, 也可以统一化。我们称这种归一化方法为”核诀窍”, 并将其应用在许多不同的数据分析任务上, 包括分类, 回归, 通过准则分析的特征提取等。

### 1.3 图小波分析

小波函数常被当做一种用于将一个函数或者信号通过不同的变换分解和表示为它的子结构的方法。小波常用在数字化的数据上, 例如通信信号或者数学函数, 也可以用在一些规律的数值结构上, 像矩阵和图像。但是图是一种随意的结构, 并且可能表示非数值化的关系, 在数据元素之间也可能存在拓扑关系。近期的一些研究证实了利用小波函数对图做多解析度分析的可行性。两个小波函数的典型就是 *Haar* 和 *Mexicanhat*。Crovello 和其他人开发了一种对于交通网络数据分析的多尺度方法。在这个应用中, 他们尝试去确定一个交通现象发生的规模。他们将交通网络表示为一个标号图, 并且用一些测量方法像每个单位时间的比特携带数来做标号。Maggioni 和其同事证实了一个通用的双正小波分析可用于图分析。在他们的方法中, 他们将利用扩散操作的二元性来激发多解析度分析。他们的方法主要适用于大空间的分类, 例如流形和图, 对于带属性的分子结构的适应性尚不清楚。在这里主要的技术问题就是如何在多解析度分析中包含节点标签。