IEEE信号处理期刊 63卷 24号 12月15日 2015

图上离散信号处理：采样定理



摘要——我们提出了一个对有向图和无向图都适用的采样定理。这一定理与传统采样理论遵循相同地规则{paradigm}。我们的结果表明，傅里叶变换下的带限图信号，是有可能实现无失真重构的。采样的信号系数形成了一个新的图信号，其对应的图结构保留了原信号的一阶差分。对一般图，我们提出了一种基于实验设计采样的最佳采样算子，以保证完美重构和对噪声的鲁棒性；对于图傅里叶变换为擦去最大鲁棒性的帧的图和Erdős-Rényi 图，随机采样使得无失真重构有很大的可能性。我们进一步建立起了有限离散时间信号处理的采样理论和以前图上信号重构的工作之间的联系。我们建立了一个基于图采样理论的图滤波器组来处理全频带图信号。最后，我们将提出的采样理论用于在线博客和数字图像的半监督分类，在这些方面，我们提出的理论相比于以前的工作，以更少的标记样本达到了相似或者更好的性能。

关键字：图上离散信号处理，采样理论，实验设计抽样，压缩传感。

# 引言

随着信息和交流爆发式增长，信号从各种源头以空前的速度产生，包括社会，引文，生物学和物理基础设施。【1，2】等。与时间序列信号和图像不同的是，这些信号具有复杂，无规律的结构，对于新的处理技术的需求引发了图信号处理这一新兴领域【3，4】。

图信号处理将经典离散信号处理拓宽到了具有复杂底层和不规则结构的信号上。这一框架通过一个图对底层结构进行建模，通过图信号对信号进行建模，并总结了从经典离散信号处理到图信号处理的概念和工具。最近的工作包括基于图的滤波器【5-7】，基于图的变换【5，8，9】，图采样和插值【10-12】，图上的不确定原理【13】，图的半监督分类【14-16】，图字典学习【17，18】，去噪【6，19】，图上社区检测和聚类【20-22】，图信号重构【23-25】，分布式算法【26，27】。

图信号处理的两种基本方式被这样认为：第一种植根于谱图理论，建立在图拉普拉斯矩阵之上【3】。由于标准图拉普拉斯矩阵被限定为对称阵和半正定阵，这一种方式只适用于具有实非负权重的无向图。第二种方式为图上离散信号处理（DSPG）【5，28】，这种方式植根于代数信号处理理论【29，30】，基于图移位算子建立，图位移算子作为基本算子，为结构给定的信号生成所有线性位移不变滤波器。图移位算子就是邻接矩阵，代表了每两个节点间的相关依赖性。图移位操作未被限定为对称操作，所以这一相应的理论框架适用于任意图，无论是有向边还是无向边，也无论是否是实信号，是否是非负权重。两个框架都使用了复杂的，不规则的结构分析信号，概括了一系列来自经典信号处理的概念和工具，例如图滤波器，图傅里叶变换，从而使基于图的应用更加多样化。

本文中，我们把经典信号处理的工作看作是图上离散信号处理框架内的采样和插值【31，32】。作为连接序列和函数的桥梁，经典采样理论告诉我们，如果采样频率足够高，一个带限信号总可以从它的采样序列实现无失真重构【33】。更普遍而言，我们可以将任何由一个线性算子导致的维度上的减少视作采样，相反地，将任何由一个线性算子导致的维度上的增加视作插值【31，34】。在这种情形下提出抽样理论就等于在高维和低维空间之间移动。

图抽样理论有着有趣的应用。例如，给出一个代表Facebook上人脉圈的图，我们可以对一部分用户进行抽样，询问它们的爱好，然后我们就可以根据重构获得所有用户的爱好。然而，图上采样这一工作没有被很好地理解【11，12】，因为图信号依托于复杂，无规律的底层结构。找到一个与采样信号系数相关联的图结构更是一个具有挑战性的工作；在Facebook这一例子中，我们对一小部分用户进行采样，一个与之相关联的图结构足以让我们推断出被采样用户之间的新的联系，甚至他们在原始的图中并未直接相连。

采样理论之前的成果【10，12，35】考虑的是在给定节点子集上唯一抽样的图信号，这种方法很难应用于有向图。它也没有解释哪种图结构支持这些采样系数。

在本文中，我们提出了一种在有向图和无向图上都支持的图信号采样理论。无失真重构对图傅里叶变换下的带限图信号来说是可行的。我们同时也提出，采样信号系数构成了一个新的图信号，这个图信号对应的图结构由原始图结构所构成。我们提出的这一采样理论遵循文献【31】的第五章，同时也符合经典采样理论。

我们将能实现无失真重构的采样算子称为合格采样算子。我们证明了对于一般图而言，为了保证无失真重构和对噪声的鲁棒性，一个基于实验设计采样的最佳采样算子被提出；对于那些图傅里叶变换为擦除具有最大鲁棒性的帧的图和Erdős-Rényi图，随机抽样有很高的概率能实现无失真重构。我们进一步建立了与有限离散时间信号处理的采样理论和图信号采样先前工作的联系。我们提出了将图信号强制变为带限信号的图滤波器组来解决全频带图信号的问题。最后，我们将提出的采样理论应用到了针对在线博客和数字图像的半监督分类中，在这种情境下，我们的理论同先前的工作相比，可以只使用更少的标记样本却拥有类似或者更好的性能。

贡献。本文最主要的贡献如下：

一个图信号采样的新体系，该体系使用线性代数中简单的工具解决复杂的采样问题。

一个新的图信号采样方法，该方法通过保留原始图信号的一阶差分实现采样。

一种新的设计图采样算子的方法。

大纲。第二章阐述了问题，并简要回顾了奠定本文基础的问题——DSPG。第三章描述了提出的图信号采样理论，和提出的采样信号系数对应的图结构的构造。第四章研究了合格采样算子，包括随机采样和实验设计采样。第五章讨论了和以前工作的联系，并将采样框架扩展到了图滤波器组的设计。第六章给出了图信号采样理论在半监督学习中的应用。第七章总结了全文，指出了未来研究方向。

# 第二章 图上离散信号处理

本章简单回顾了图上离散信号处理相关概念；完整介绍在文献【4，28】中可以找到。它是一个这样的的理论框架，该框架将经典离散信号处理从线或者矩形格这样的规则域推广到通常由图描述的不规则结构。

A. 图移位

图上离散信号处理利用由图表示的复杂的，不规则的结构研究信号，其中是节点的集合，代表图移位，或者说带权重的邻接矩阵。它代表了图*G*的关联性，这里的图既可以是有向图也可以是无向图（值得注意的是，标准图拉普拉斯矩阵之恶能表示无向图【3】）。连接节点和的边的权重是一个代表第*n*个和第*m*个节点之间的潜在关系的定量的表示，这种关系例如相似性，独立性或者沟通模式。为了保证移位算子缩放合适，我们把图移位算子归一化成满足。

B. 图信号

给定图的表示，图信号定义为在图节点上被分配了信号系数的图。一旦节点顺序固定，图信号就可以被写作一个向量

|  |  |  |
| --- | --- | --- |
|  |  | 式(2-1) |

这里第*n*个信号系数对应节点。

C. 图傅里叶变换

一般来说，傅里叶变换对应使用不随滤波改变的基本元素进行的信号拓展；在这里，这个基本元素就是图移位矩阵A的特征基（或者说，如果A的特征基不存在，就是A的Jordan特征基）。简单而言，假设A有完整的特征基，那么A的频谱分解为【31】

|  |  |  |
| --- | --- | --- |
|  |  | 式(2-2) |

矩阵*V*由A的特征向量组成，是对应特征值的对角矩阵。这些特征值代表图上的频率【28】。这里我们先不指定图频率的顺序，稍后将解释原因。

定义1：的图傅里叶变换为

|  |  |  |
| --- | --- | --- |
|  |  | 式(2-3) |

逆傅里叶变换为

|  |  |  |
| --- | --- | --- |
|  |  | 式(2-4) |

式(2-3)中的向量表示信号在特征向量基的展开，也描述了图信号的频率成分。逆傅里叶变换从图信号的频率内容通过连接图频率分量来重构图信号，这里的图频率分量由图傅里叶变换的系数决定。

表1 文中使用的关键符号

|  |  |  |
| --- | --- | --- |
| **Symbo** | **Descript** | **Dimension** |
|  | 图移位矩阵 |  |
|  | 图信号 |  |
|  | 图傅里叶变换 |  |
|  | 频域图信号 |  |
|  | 采样算子 |  |
|  | 插值算子 |  |
|  | 采样指数 |  |
|  | 的采样系数 |  |
|  | 的第K个系数 |  |
|  | V的第K列 |  |

# 第三章 图采样

之前的研究中，图信号的采样理论基于谱图理论【12】。图信号的带宽是基于图频率的值来定义的，图频率的值对应图拉普拉斯矩阵的特征值。由于每个图都有自己的图频率，实践中很难指定一个一般的图截止频率；同时，计算图频率的所有值的计算效率很低，特别是对规模较大的图。

本章我们提出了一个新的图信号采样框架。这里带宽的定义是基于信号在图傅里叶变换域的非零系数的个数的。因为图傅里叶域的每个信号系数都对应一个图频率，所以带宽的定义也可以基于图频率的个数。这使得我们提出的采样框架和线性代数有了紧密的联系，即我们可以使用线性代数中的简单的工具来表示在复杂的，无规律的图上的采样。

A. 采样和重构

假设我们要对图信号进行点采样得到采样信号，这里表示采样指数的序列，且。然后我们对进行插值可以得到，这样就可以准确或近似还原信号。采样算子是一个从到的线性映射。定义为

|  |  |  |
| --- | --- | --- |
|  |  | 式(3-1) |

插值算子是一个从到的线性映射（见图3.1）

|  |
| --- |
|  |
| 图3.1 先采样后插值 |

可以得到采样和插值操作为

|  |
| --- |
| 采样：，  插值： |

这里可以准确或近似恢复信号。考虑两个采样策略：随机抽样意味着采样指数是从中独立随机选择的；实验设计抽样意味着采样指数可以是预先选择好的。很显然随机抽样是实验设计抽样的一个子集。

当是单位矩阵时，对所有的信号都可以实现无失真重构。一般来说这是不可能的因为。然而，对于具有特定结构的信号，无失真重构是可能的，我们把这种信号定义为带限图信号，就像在经典离散信号处理中的一样。

B. 图信号采样理论

现在，我们定义一类带限图信号，它使得无失真重构变为可能。

定义2：如果存在一个，对于图信号使得它的傅里叶变换满足

|  |  |  |
| --- | --- | --- |
|  |  | 式(3-2) |

我们就称图信号是带限的。最小的满足条件的*K*叫做图信号的带宽。非带限图信号又被叫做全频带图信号。

注意到这里的带限图信号不一定意味着低通或光滑。因为我们没有指定频率的顺序，我们可以在图傅里叶变换矩阵中对特征值重排并置换对应的特征向量，从而在图傅里叶变换域选择出任意波段。只有当特征值按下降的顺序排列时，带限图信号才是光滑的。带宽的限制在这里相当于限制已知的图傅里叶变换域中非零信号系数的个数。这个概括对于表示非光滑图信号可能很有用。

定义3：在内，带宽最大是K的图信号的集合是一个封闭子空间，表示为，其中与式(2-2)中相同。

定义带宽的时候，我们专注于图频率的个数，之前的工作【12】关注点在图频率的值。使用图频率的值有两个弊端：(a)考虑图频率的值时，我们忽略了图的离散性；因为图频率是离散的，同一图上的两个截断图频率可能会对应相同地带限空间。例如，假设一个图的图频率为0，0.1，0.4，0.6和2；无论我们将阶段频率设置为0.2还是0.3，都会形成相同地带限空间。(b)图频率值不能在不同的图之间进行比较。因为每个图都有自己的图频率，两个图上相同地截断图频率值可能表示不同的结果。例如，一个图的图频率是0，0.1，0.2，0.4和2，另一个图的图频率是0，1.1，1.6，1.8和2；当我们把截止图频率设置为1时，也就是说，我们保留所有不大于1的图频率，第一个图保留了四分之三的图频率，第二个图只保留了四分之一的图频率。因此，图频率的值没有必然地给出一个关于带限空间直观的理解。使用图频率的个数的另一个关键优势在于建立了一个同线性代数的紧密联系，这使得我们可以在带限图信号的采样和插值时可以使用线性代数中的简单工具。

在文献【31】的定理5.2中，作者通过投影展示了向量的重构，这为经典采样理论奠定了理论基础。按着这一定理，我们可以得到如下的结果，其证明可以在【34】找到。

定理1：使满足

|  |  |  |
| --- | --- | --- |
|  |  | 式(3-3) |

这里表示V的第K列。对所有的，无失真重构通过选择如下的决定

|  |  |  |
| --- | --- | --- |
|  |  | 式(3-4) |

上式满足是一个的单位矩阵。

在已知条件的情况下，定理1对所有图傅里叶变换域有非零元素的图信号都适用。

类似于经典采样定理，图信号的采样速率也有一个下限，即采样点数应不小于带宽。当时，，因此，不可能是一个单位矩阵。当的时候，若要是一个单位矩阵，应当是矩阵的逆；当时它是矩阵的伪逆，其中的冗余对于降低噪声的影响会很有用。为简单起见，我们只考虑和可逆的情况。当时，我们只需根据M个采样信号系数来选择K从而保证采样点大小和带宽相同。

从定理1中，我们可以知道即使对带限图信号，一个随机采样算子也可能不会实现无失真重构。当采样算子满足全秩假设式(3-5)时，我们把它称为合格采样算子。为了满足式(3-5)，采样算子应该在中选择一组K线性独立的行。因为V是可逆的，所以V中的列向量线性无关，且恒成立；换言之，在中总存在一组K线性无关的行。因为图移位矩阵A已经给出，所以可以独立于图信号找到这样的集合。给定这样的集合后，定理1保证带限图信号的无失真重构。有些高效的算法可以找到矩阵中线性无关的行，比如QR分解；详情在文献【36，31】。因为我们只需要知道图结构就能设计一个合格采样算子，这就遵循了实验设计采样。我们将在第四章展开讲这个话题。

C. 采样后的图信号

我们只证明了图信号在带限的情况下无失真重构是可能的。现在证明采样后的信号系数构成了一个新的图信号，它对应的图移位可以由原始图移位构成。

虽然接下来的结果可以很容易概括为，为了简单起见，我们只考虑的情况。使采样算子和插值算子符合定理1中的情况。对所有的，可以知道

|  |  |  |
| --- | --- | --- |
|  |  |  |

这里的代表的第K个系数，（a）由定理1得出，（b）由定义2得出。因此我们可以知道

|  |  |  |
| --- | --- | --- |
|  |  |  |

由此可知，采样后信号系数和频率分量形成了一个傅里叶变换对，因为可以从通过构造，也可以由通过构造。这说明，根据定义1和谱分解式(2-2)，是一个与图傅里叶变换矩阵U相关联的图信号，也是一个新的图移位

|  |  |  |
| --- | --- | --- |
|  |  |  |

这里有

|  |  |  |
| --- | --- | --- |
|  |  |  |

其中是对进行采样的前K个特征值组成的对角矩阵。因此可以得到如下的定理。

定理2：令，则它的采样信号为，其中是合格采样算子。然后，和图信号对应的图移位矩阵为

|  |  |  |
| --- | --- | --- |
|  |  | 式(3-5) |

其中。

从定理2中我们可以知道，图移位矩阵是由对特征向量矩阵的行采样和对原始图移位矩阵A的前K个特征值采样共同构成的。我们通常会说是从A中采样得到的，并在图傅里叶变换域保留了一定的信息。

因为x的带宽是K，所以频域的前K个系数为，另外个系数；换言之，对原信号x和采样信号进行傅里叶变换后可以知道，它们的频率分量是相同的。

类似于定理1，在图傅里叶矩阵上对特征值重新排列，对相应的特征向量进行置换，定理2适用于所有在图谱域具有有限支持的图信号。

D. 采样图信号的性能

我们认为，是支持采样信号系数的图移位，它来自于采样图信号的图傅里叶变换和图移位之间的数学等价性。实际上，我们含蓄地提出了一种采样图的方法。由于采样图总是丢失信息，我们现在研究哪些信息保留。

定理3：对于所有的，有。

证明：

|  |
| --- |
|  |

这里最后一个等号从得知。

这一项测量原始图信号与其移位后的差异。这也被称为x的一阶差分，而这一项测量的是采样指数的一阶差分。此外，是基于范数的图的总变化，这是衡量图信号光滑度的量化特征【28】。当使用采样图来表示采样信号系数时，我们会丢失采样节点与所有其他节点之间的连通性信息；尽管如此，他仍保留了采样指数的一阶差。我们并没有像之前的工作【37】一样专注于保持连通性，而是强调信号和结构之间的相互作用。

E. 举例

考虑一个五节点的有向图，其图移位矩阵为

|  |
| --- |
|  |

与之对应的逆傅里叶变换矩阵为

|  |
| --- |
|  |

该信号的频率为

|  |
| --- |
|  |

令K=3，生成一个带限图信号如

|  |
| --- |
|  |

并且原信号满足

|  |
| --- |
|  |

此时x的一阶差分为

|  |
| --- |
|  |

我们可以检查V的前三列，可以看出三行的所有集合都是独立的。根据抽样定理，我们可以通过采样任意三个系数对x无失真重构；例如，对第一，第二，第四个系数进行采样。然后得到

|  |
| --- |
|  |

并且采样算子是合格采样算子

|  |
| --- |
|  |

我们通过如下的插值算子重构信号x

|  |
| --- |
|  |

重构的过程如图3.2所示

|  |
| --- |
|  |
| 图3.2 先采样后插值。箭头表示边是有向的 |

采样信号的逆傅里叶变换矩阵为

|  |
| --- |
|  |

采样频率为

|  |
| --- |
|  |

采样图移位可以写作如下

|  |
| --- |
|  |

的一阶差分为

|  |
| --- |
|  |

我们可以看到，采样图移位包含自循环和负权重，这似乎与矩阵A有所不同，但保留了A的部分频率内容，因为是根据V进行采样的，是从A中采样的。也保留了x的一阶差分，这很好地验证了定理3.

# 第四章 合格算子采样

正如第三章所示，