

11/18 思路报告

杨振东，温晋，杨晓蓉

关于多边形表示方法

在基于散点的展示方法中，当数据量提升而尺度被限制，地图上将不可避免地产生重叠。

因此，我们难以找到一个合适的点绘制方法使得地图的呈现效果同时满足：

1. 较少的重叠
2. 较均匀的像素分布

因此，认为**以面代点**是一种有效的可视化分析手段。同时，各种多边形表示方法（如 SuperPixel 和 Voronoi）并没有本质的区别。

其**合理性**可按照如下方法验证：

通过插值法为原始数据生成预测的值分布；对于我们的生成图像，对比其**每一个像素**所表征的属性值与其对应位置的插值结果，求差的绝对值或平方并汇总取平均。结果越小，证明对于生成图像的每一个像素，更能够真实地反映它的值，置信度越高。此目标亦可用于多种多边形表示方法的横向对比。**生成图像更能够反映原始插值的方法将被考虑。**

这个指标可用于衡量生成图像的表征准确性（已获得陈老师的确认）。

综上，可定义损失函数 Loss1：

$$Loss1 = \sum_{i,j} (val_display_{i,j} - val_interpolate_{i,j})^2 / count$$

关于是否可以使用 Loss1 指导多边形的生成，目前暂未占到合适的解决方案。

自此，可以提出三个目标：

1. **生成图像不产生过度绘制和重叠；**
2. **像素区域能更大程度上连续；**
3. **生成图像的每一个像素都能够尽量准确地还原原始数据分布。**

关于采样

Voronoi 图是一种已经在响应领域产生应用的成熟的方法。但是基于上条的讨论，由于其生成方法限制其每一个原始数据点都直接对应到一个多边形，直接使用 Voronoi 图将会**继承原始数据的视觉重叠，无法达成目标1**。因此，周老师提出在生成之前预先进行采样的方式。

有关采样的构思，现在此扩展。

以采样简化底层多边形结构

按照上述分析，为了避免视觉重叠，将一些密集区域的点用一个单独的点代替表示将会更好，也更有利于生成多边形的视觉优化。但同时，为了维持其属性值，主要思路分为以下两条：

1. 在**采样中**考虑进数据点的**属性值**，以获得期待的空间和属性分布；

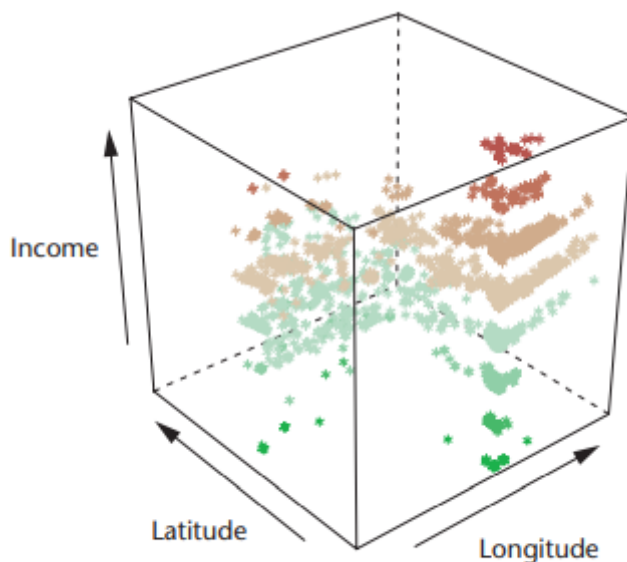
2. 在**采样后**，将生成多边形按照其区域内所对应的**全部原始数据点**的总体特征表示。

对于采样算法的思考

本部分内容曾经由我与张欣隆学长讨论过，但未能得出深入结论。在会议上听得周老师的想法之后想到简单完善一下，也在此写下作为补充。

蓝噪声采样在图形方向应用广泛，主要是因为其**泊松盘**的设计，使得点能够在空间上基于密度间隔分布。

但是，属性值本身其实也是一个维度，可以与空间坐标结合构成三维空间——



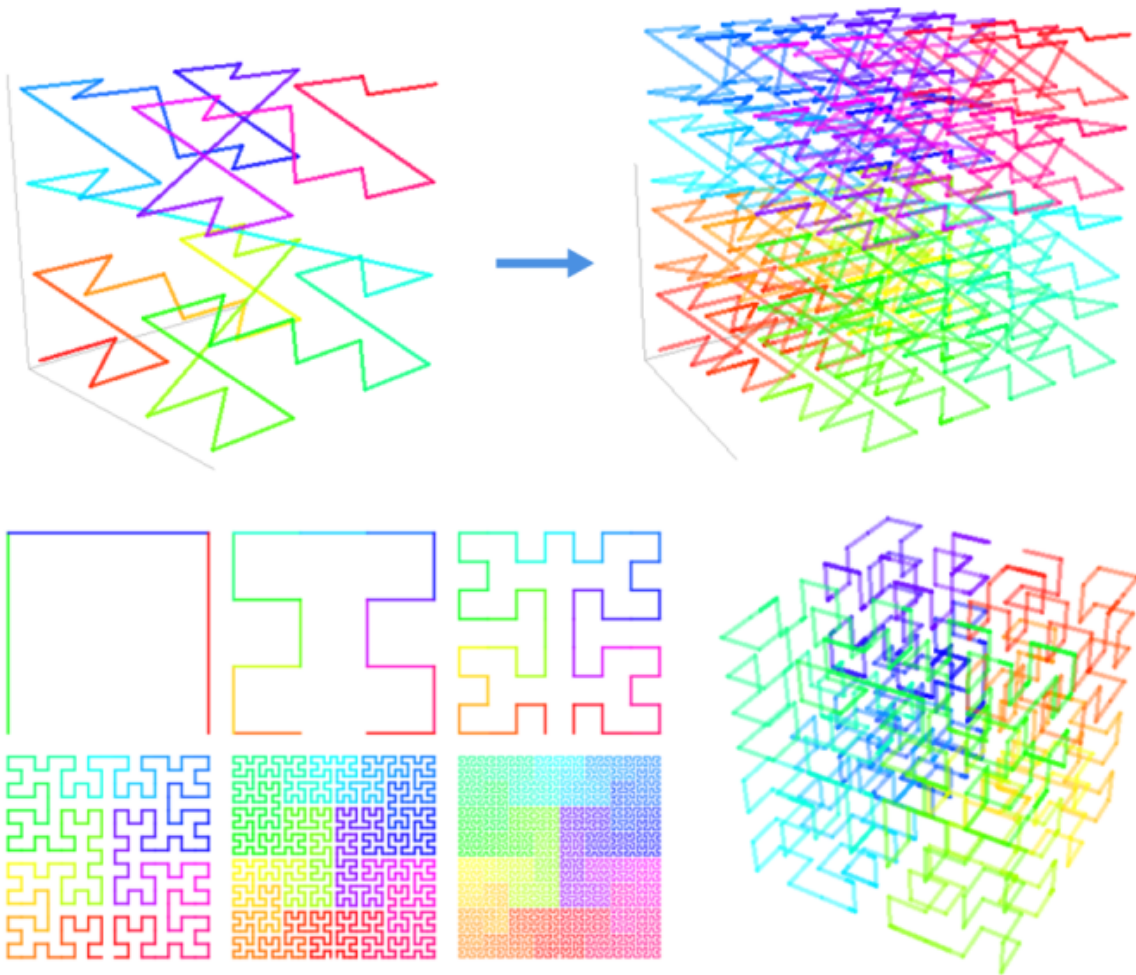
(a)

(配图来自 Daniel A. Keim 于 2004 年发表的论文 *Visual Data Mining in Large Geospatial Point Sets* 。他在其中针对固定像素大小的地图上的散点堆叠问题，提出了考虑属性值剖分的数据抖动。)

如果将**蓝噪声采样及其泊松盘的操作扩展到三维空间**，理论上可以同时考虑到空间位置和属性值的差异进行采样，获得更有代表性的样本点。当然，这里有一个比较棘手的问题：是否有计算三维空间核密度的成熟方法。同时，将属性值在其对应的维度上等比例地放大或缩小，可视为对**属性值 vs 空间距离**一对因素的**权重分配调整**。当属性值的极差越大，相邻空间位置内的若干数据点的属性值分布的层次将越多；当属性值的极差越小直至趋近 0，采样将更高地考虑空间距离的差异，到 0 时即为原始的不考虑属性值的蓝噪声采样。

同时，我们也并非只能使用蓝噪声。可以预见的是，蓝噪声能够取得很好的分布，但它的性能消耗和复杂度将会是一个很高的数字。

为了维持地理空间分布，我们曾经使用过 Z-order 指导采样。作为理论上能够穿过空间中任意点的曲线，Z-order 使用的 Z 阶曲线（又：Peano 曲线）可以达到编码再降维的作用，以更简单地获取到数据点集的划分。同样，具有相同性质的 Hilbert 曲线也能实现该目标。而值得注意的是，这两种空间填充曲线都可以延伸至三维空间。



以上，是否能够通过**直接将属性值引入维度空间进行采样**达到同时考虑空间分布保持和属性值分布保持的目标，私以为是一个可以考虑的点。

总结：工作流程

可能使用采样等预处理手段，生成底层的多边形，再在此基础上进行合并、层次聚类等操作，达到增强视觉效果、提高可视化的信息表达、同时较大限度上表征原始数据的属性分布。