

# PKU Visualization Blog

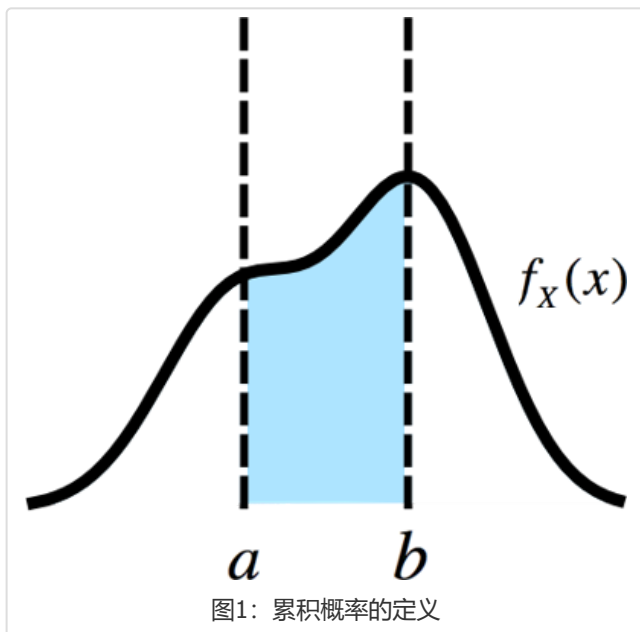
北京大学可视化与可视分析博客

## 区间似然树：一种可视探索不确定性数据集的紧凑高效的表示方法 (Range Likelihood Tree: A Compact and Effective Representation for Visual Exploration of Uncertain Data Sets)

作者: Fan Hong 日期: 2017年5月27日

本文针对的数据是具有不确定性的标量场数据。我们知道，普通的标量场数据在一个空间格点上只会确定的取一个数值。而在具有不确定性的标量场数据中，一个空间格点会以一定概率分布取不同的标量值。因此，这种数据也被称为概率分布场数据。这类数据在原有的空间维度上新增加了一个分布维度，因此对这类数据的可视化和分析会十分困难。已有的方法或者基于一些统计指标对概率分布进行约减，或者通过定义概率分布间的相似性来进行分析。而本文则提出使用累积概率来进行分析。

对于一个概率分布来说，给定一个区间 $\Gamma=(a, b)$ ，其在这个区间的累积概率就是概率分布函数在这个区间的定积分。通过使用累积概率，一个概率分布就被约减为单个标量值。而对于具有不确定性的标量场数据来说，给定一个区间，在每个空间格点上都求出对应的累积概率，那么整个概率分布场就转化成为了一个标量场。我们称之为区间似然场 (range likelihood field, RLF)。直观地理解，这个场就是表示每个空间格点有多大概率取值在给定区间中，从而可以帮助提取具有相近累积概率的空间位置。对于区间似然场的可视化，可以直接采用针对体数据的等值面可视化、体渲染等技术。



更进一步地，通过选择多个区间，可以得到多个区间似然场，从而将整个概率分布场就转化成为了多标量场数据进行可视化与分析。多个区间的选择，也能更加全面的描述概率分布，并基于此提取相似的概率分布。图2中就列举了选择两个区间时三种可能的概率分布情形。

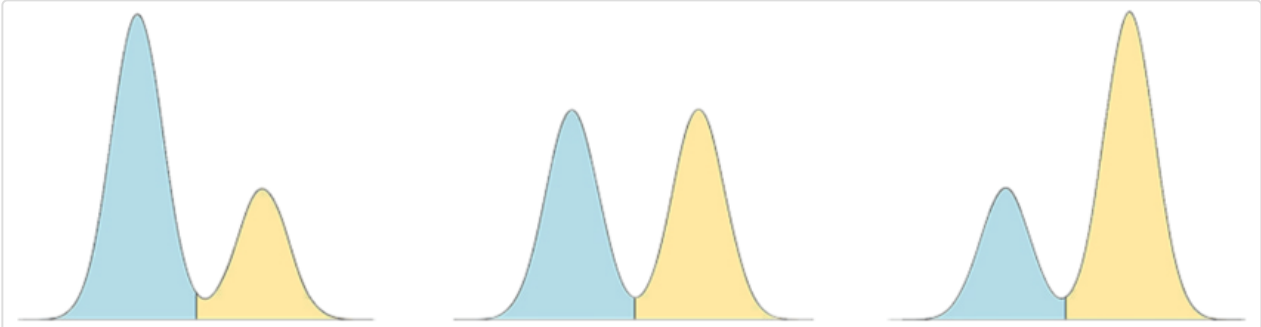


图2：选择2个值区间时，3种不同的概率分布情形

现在，如何选择值区间是一个关键问题。作者提出了一种自底向上的工作方式，如图3所示：整个值域被划分成若干个极小的值区间，对于每个值区间生成对应的区间似然场；然后计算两两区间似然场的相似度（距离）；最后基于距离矩阵，使用凝聚层次聚类 (agglomerative hierarchical clustering) 的方法将区间似然场组成二叉树的形式。其中，对于区间似然场的距离定义，可以把每个场进行归一化之后看作一个空间分布函数，然后区间似然场的距离就转化为两个空间分布函数的Jensen-Shannon散度 (JSD)。基于这个距离定义，在凝聚层次聚类中，初始时每个区间单独为一个聚类，每次选择两个距离最小的聚类进行合并，直到所有区间合为一个，即整个值域。整个合并过程可以用一颗二叉树来表示，称之为区间似然树 (range likelihood tree, RLT)，其中每个节点就对应一个区间，叶子区间对应初始的值区间，而根节点对应最终的值域。

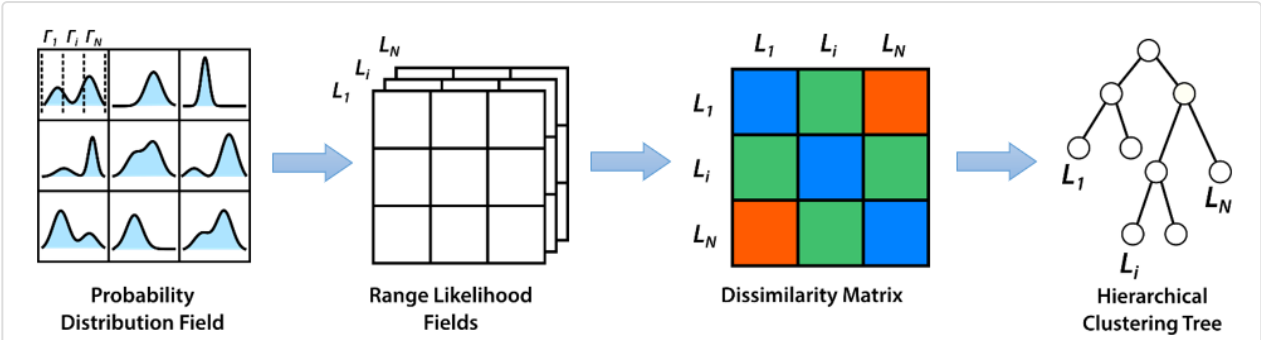


图3：算法流程

作者基于此方法提出了一套可视探索框架，包含RTL视图、RLF视图、多标量场传递函数部件、以及多标量场分类视图。图4展示了RTL视图，其中区间似然树用环状布局来展示，在树结构外围用一圈圆环表示整个值域。选择一个节点之后，圆环上对应区间将用蓝色高亮。在最外围则是整个概率分布的密度图可视化，即将各个格点位置的分布曲线叠加起来，再转换为密度图的形式展现。用户选择一个节点之后，对应的区间似然场用体绘制形式直接展示出来。

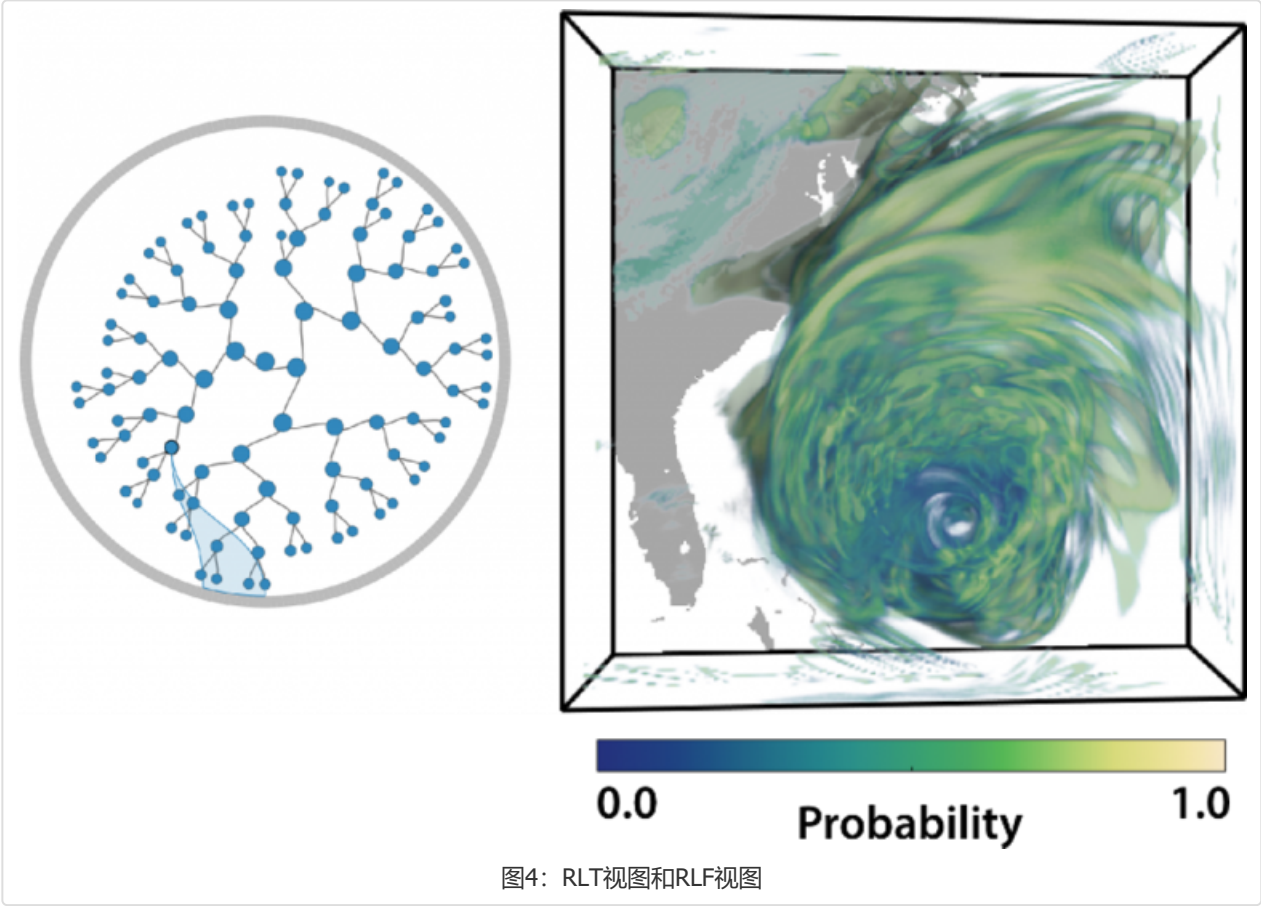


图4：RLT视图和RLF视图

用户也可以选择多个区间生成平行坐标：每个轴对应一个值区间的累积概率取值，每条折线则对应一个空间格点，将其在各个区间上的累积概率连起来。与高维数据中的平行坐标类似，用户可以在轴上选择一段区间，从而选择一个空间格点的子集。用户可以对每次选择的区间格点设置不同的颜色和透明度，然后在空间视图中观察它们的分布，如图5所示。

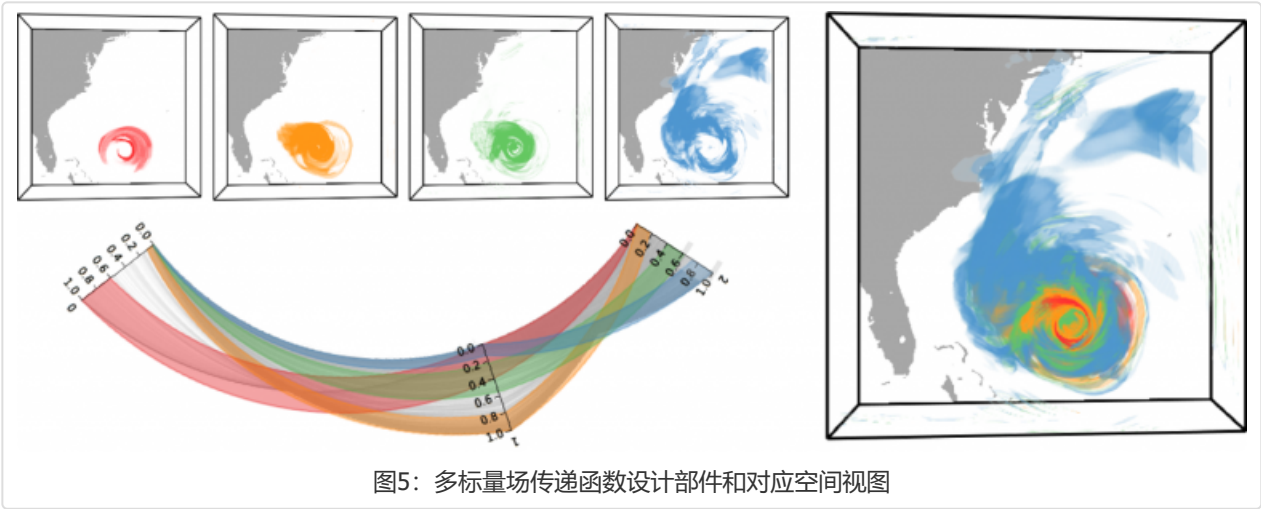


图5：多标量场传递函数设计部件和对应空间视图

作者在Massachusetts Bay集合模拟数据中试验了其可视探索框架，使用的变量是叶绿素a。作者在初始的RLT视图中选择了5个区间进行了探索，如图6所示。对于其中的三个区间，用户进一步采用平行坐标的形式进行探索，从中选择了4个不同的特征进行可视化，如图7所示。

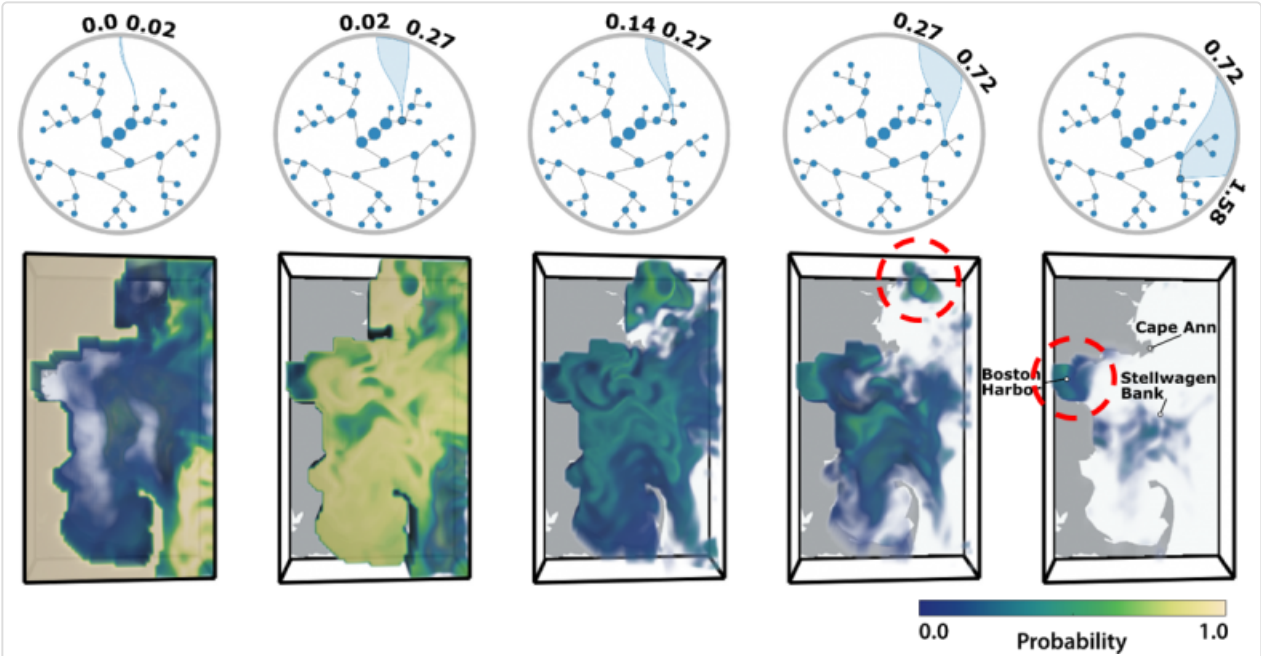


图6：Massachusetts Bay集合模拟数据的RLT视图和5个值区间的RLF渲染结果

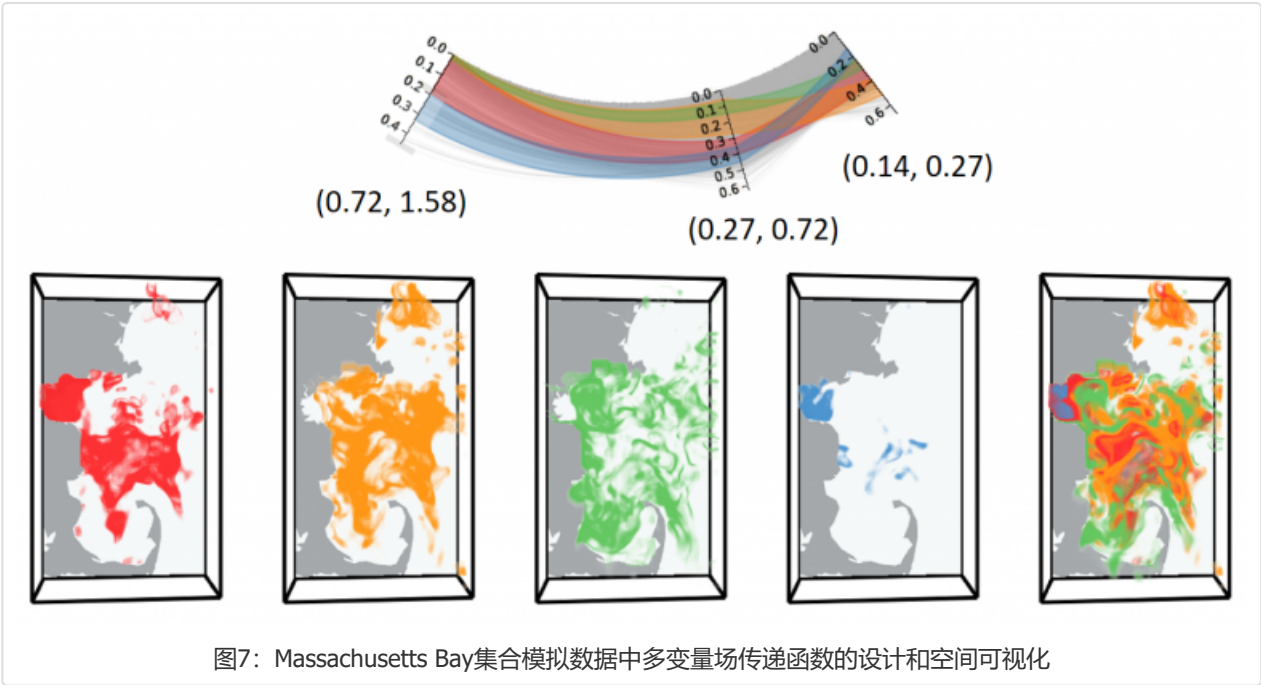


图7：Massachusetts Bay集合模拟数据中多变量场传递函数的设计和空间可视化

总结起来，这篇工作基于累积概率从值区间的维度来对概率分布场进行约减，是一种十分新的方式，并且取得了不错的效果。

论文报告

- ← Squares: 支持多类别分类器的交互式分析展示  
(Squares: Supporting Interactive Performance Analysis for Multiclass Classifiers)
- 虚拟牵开器：使用基于物理变形的交互式数据探索系统  
(Virtual Retractor: An Interactive Data Exploration System Using Physically Based Deformation) →

评论关闭。

RSS 订阅

功能