

PKU Visualization Blog

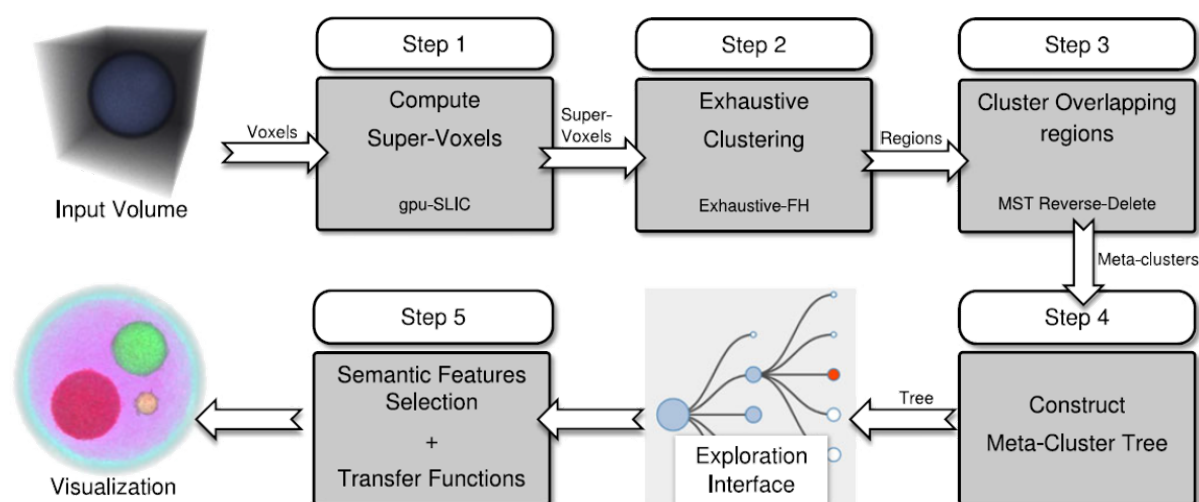
北京大学可视化与可视分析博客

特征乐高：一种使用穷举聚类进行超体素聚类的体数据探索方式 (FeatureLego: Volume Exploration Using Exhaustive Clustering of Super-Voxels)

作者: Yang, Changhe 日期: 2019年10月27日

对于体数据的探索十分困难。一方面，体数据中蕴含着丰富的特征，探索过程中容易有所遗漏。另一方面，体数据中又充斥着噪声和用户不感兴趣的区域，需要精心进行特征的提取和筛选。对于体数据进行基于特征选择的探索方式通常基于体素聚类。在以往的工作中，交互式地聚类方式，需要用户对于聚类边界进行细致的调整，给用户带了巨大的负担。而由此衍生的多重聚类 (Multiple Clustering)，则通过对于聚类算法超参数的采样，获取尽可能多的聚类结果。而在此过程中，需要大量的试错式的尝试，才能得到良好的参数采样才能得到最终良好的聚类结果，同时由于采样的问题，特征的提取可能不够穷尽而导致有所遗漏。本文提出了一种基于穷举聚类的体素聚类方式，其命名而“特征乐高”，因为其在探索过程中能够提供给用户穷尽的特征聚类结果，就像基本的积木一样，用户可以有效的进行特征选择，加以组合，最终生成定制化的体数据可视化结果。

本文提出的方法主要分为两个部分，即聚类特征提取算法和交互交互式选择，其基本的流程图如下。



聚类特征提取算法包含三个层次的聚类方法。第一阶段是将体素聚类成超体素。超体素指具有相似的强度值分布的体素块，是一种更加紧致的数据表达形式。该阶段采用的算法为三维简单线性迭代聚类 (3D SLIC)，该算法为二维图像处理领域的算法拓展。其原理类似于K均值聚类算法，但每个聚类中仅仅辐射周围较小的领域，因此在初始化时，增添了均匀分布聚类中心和局部微调以免陷入边界的限制改进。

第二阶段则是将超体素聚类成区域。该阶段采用的算法为改进的FH聚类算法。其同样为图像处理领域FH图像聚类算法的拓展，并加入了穷举聚类的改进。基本的三维FH聚类算法思想类似于自底而上的层次聚类方式，同时采用了图聚类思想。其将每一个超体素视为一个节点，两两连接形成一条边，边权为超体素之间的不相似度，采用卡方距离度量超体素内部强度分布的差异。算法初始将所有节点视为单个的区域，通过权重逐

增的方式依次考虑每一条连接两个区域的边，如果其满足以下不等式，则该边崩溃两个区域融合。该公式描述了链接两个区域的边能否在构建基于两区域的图的最小生成树过程中，能否先于两区域独立完成最小生成树的构建。同时该公式增添了一个带有超参数的松弛项，以保证初始阶段，区域的融合可以进行。

$$w(C_1, C_2) \leq \min \left(\text{Int}(C_1) + \tau(C_1), \text{Int}(C_2) + \tau(C_2) \right).$$

Minimal edge weight within the edges connecting two regions

Minimal edge weight in the minimum spanning tree of region

Parameter k dividing the size of region

该超参数可以有力地影响算法聚类的最终结果，想要得到尽可能多样的聚类结果，需要选取不同的超参数本文观察到该算法具有单调性、递归推导性和结果的一致性，发现在当超参数的值于一定的区间内，顺序考虑每一个链接两个区域的边是否崩溃导致区域融合时，所有的决策均可相同，因此最终能够得到完全相同的聚类结果。具体来说其通过维护一个区间（初始为0到正无穷），并将左端点作为聚类算法超参数进行聚类。当算法执行过程中一条边崩溃导致区域融合时，任何大于区间左端点的值均可以导致该边崩溃，因此在此区间内所有超参数均可以保证结果相同；当算法执行过程中一条边未满足崩溃的区域融合时，可以通过更新区间的右端点，来保证当超参数取该区间内所有的值时，在该处执行同样的决策。更新的值可以通过上述区域融合条件的不等式取等号时来计算。该算法不断迭代这个更新的过程，直到计算完成，得到该参数下的聚类结果。此时，该算法给出了一个超参数取值区间，可以保证相同的聚类结果。此时考虑下一轮计算，初始维护的区间更新为上一轮区间右端点至正无穷，不断迭代计算直到所有的超体素均被聚类为同一个区域，算法结束。该算法的伪代码如下。

Algorithm 1. Exhaustive FH Clustering

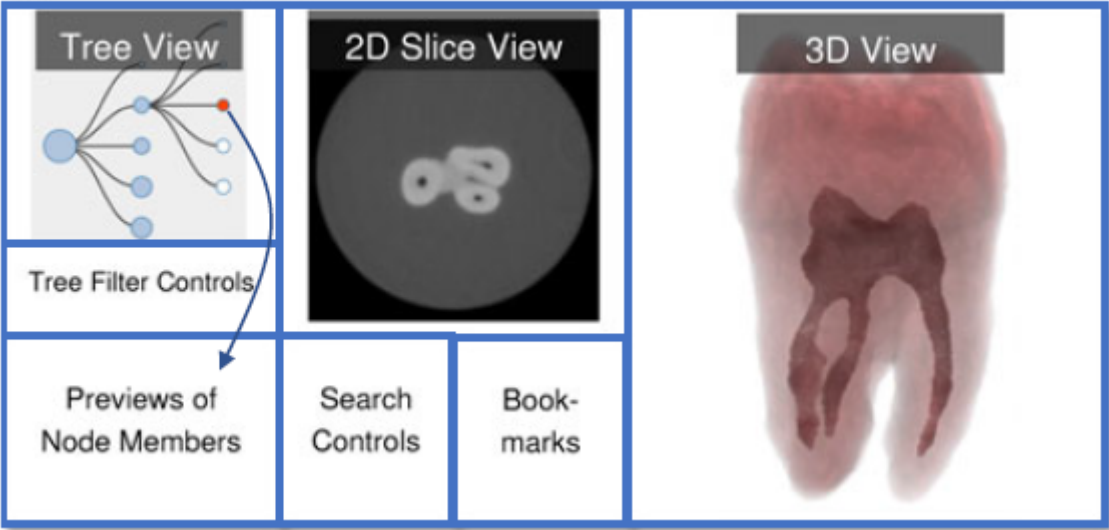
```

Construct compact super-voxels  $B$ ;
Construct adjacency graph  $G(V, E)$  for super-voxels;
Sort  $E$  by non-decreasing order of edge-weights;
Initialize  $k \leftarrow \{0, \infty\}$ ;
do
     $S \leftarrow B$ ;
    for each  $e : E$  do
        if  $e.weight \leq k[0]$  then
            Merge regions connected by  $e$  in  $S$ ;
        else
            Calculate  $k_e$  using Eq. (2);
            if  $k_e < k[1]$  then
                 $k[1] \leftarrow k_e$ ;
    OutputList.insert(  $S$  );
     $k \leftarrow \{k[1], \infty\}$ ;
while  $S.RegionCount > 1$ ;
return OutputList;

```

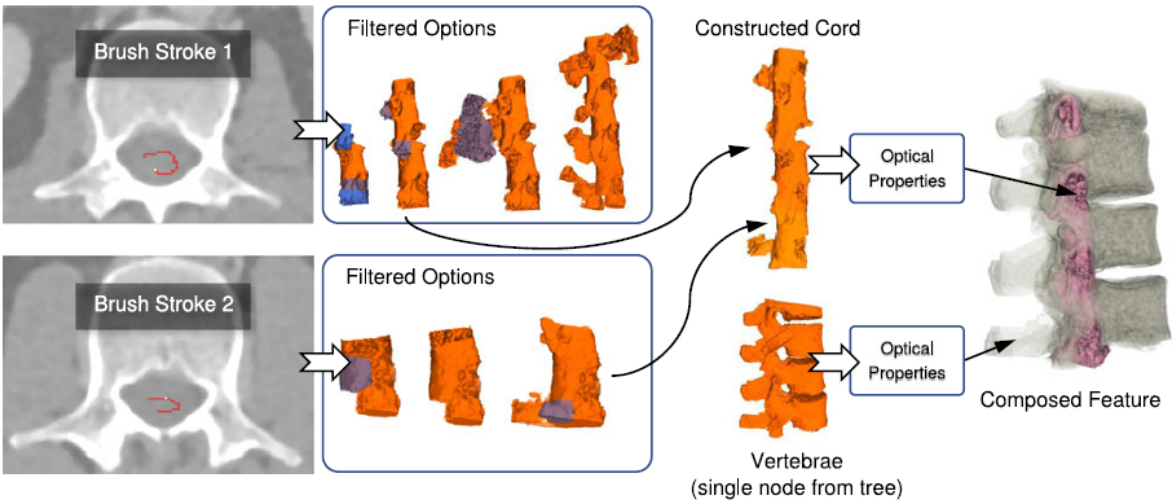
第三阶段则是将区域聚类成超聚类（meta-cluster），该过程中生成的聚类的区域成员之间在原始体数据上有十分巨大的重叠关系。该阶段同样采用图聚类的方式。通过将上一阶段计算的所有超参数取值可能下的穷尽的区域聚类的结果中的每一个区域定义为一个节点，用杰卡德距离来定义节点相连的边之间的权重，来描述不同区域间的重叠程度关系。通过将这样的一个图建立最小生成树，并依据用户选定的参数删除最小生成树种所有权重大于该权重的边，最终得到的各个连接子图则各自为一个超聚类。

基于以上的特征提取聚类算法，该方法可以得到一系列的特征候选超聚类，特征乐高同样提供了对于用户进行交互式特征选择的帮助。下图为该方法中提出的用户特征选择和可视化面板。

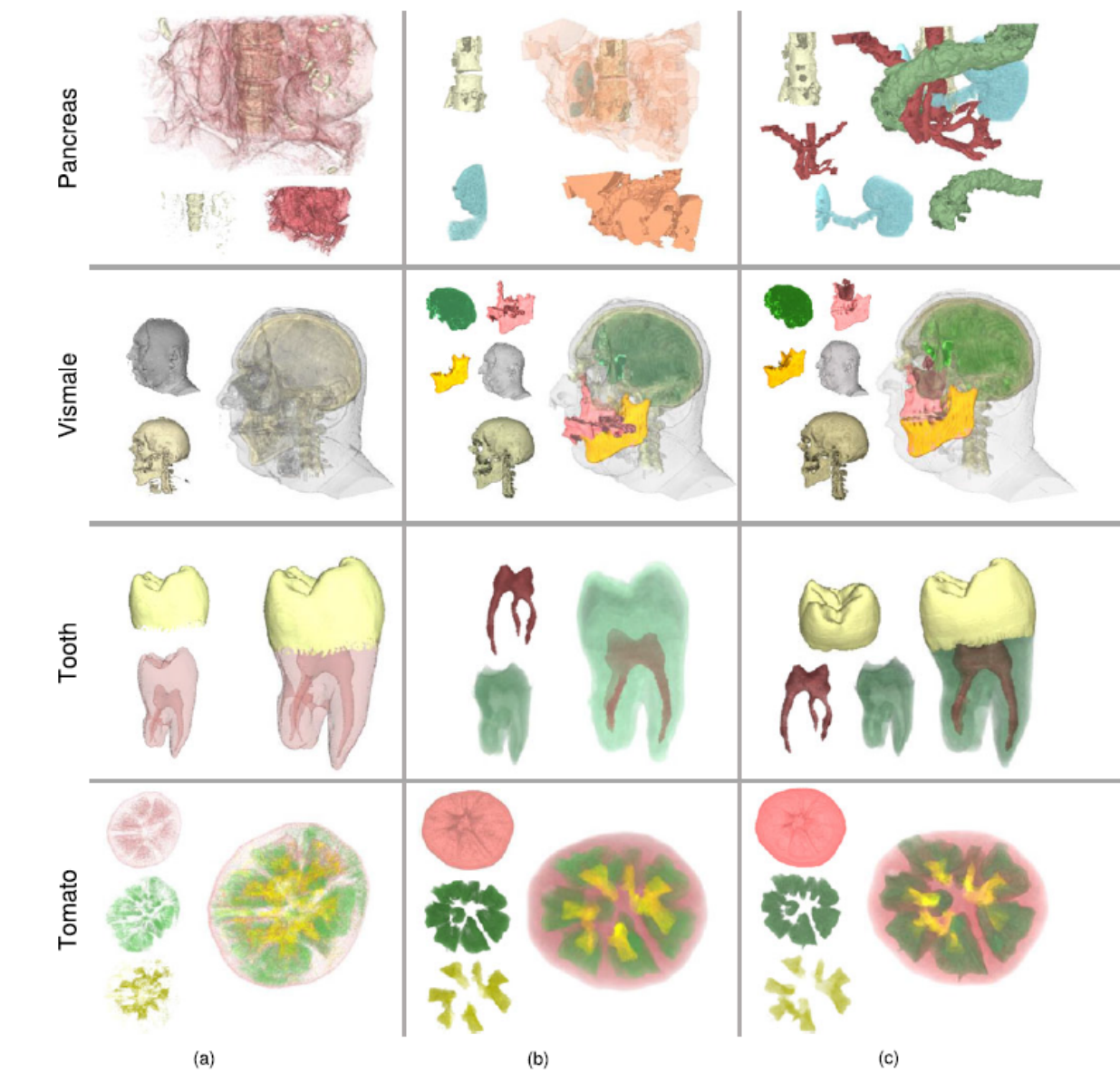


首先其建立了一个树形的数据结构来保存得到的各个超聚类。超聚类树的每一个节点是一个超聚类，树结构中一个节点的父节点定义为该超聚类节点所占区域的容量最小的超集。由于该定义下，可能会出现祖先节点对于后代节点包含关系的不完整，该方法通过复制冗余节点的方式来实现修正。兄弟节点的顺序则通过其容量的大小。这可以为用户的筛选、查询提供便利。总的来说，树形结构使得用户可以通过点选节点的方式，选择超聚类并展开最终获得其期望获得的特征区域。

该方法同样提供了筛选和查询的功能，保证用户能有效地选择其感兴趣的特征。用户可以在切片可视化视图中刷选一定的体素，算法结合用户的筛选条件在超聚类树结构上自底而上地搜索，并返回候选的节点，用户可以进一步地选择其所需要的特征区域，并通过调节传递函数生成定制的可视化结果。



本文最后对于多种方法进行了详尽的效果比较，可以看出该方法在提取特征的完整性和特征结构保留的健全性上都有较高的优势。



总的来说，本文提出了一种完备的针对体数据的特征提取和选择的新颖探索方式，通过改进已有的聚类算法并引入穷举聚类的思想，辅之以高效的数据管理方式，能够为用户提供简单、快捷的交互体验，值得一读。

参考文献

[1] Shreeraj J , Saad N , Kaufman A E . FeatureLego: Volume Exploration Using Exhaustive Clustering of Super-Voxels[J]. IEEE Transactions on Visualization and Computer Graphics, 2018:1-1.

论文报告 体数据可视化, 特征分析, 科学可视化

← 用于探索，监视和预测道路交通拥堵的可视化分析系统
(A Visual Analytics System for Exploring, Monitoring, and Forecasting Road Traffic Congestion)

多视图多类别散点图的利用集合覆盖优化的数据采样
(Data Sampling in Multi-view and Multi-class Scatterplots via Set Cover Optimization) →

评论关闭。

功能

登录

文章RSS

评论RSS

WordPress.org

链接

北京大学可视化与可视分析研究小组主页 – PKU Vis Home Page

北京大学可视化研究维基 – PKU Vis WIKI

分类目录

应用

新闻

未分类

活动

研究

论文报告

标签

ChinaVis graph interaction PacificVis

pacificvis2019 pvis2016 不确定性 主题模型 交互

交互设计 人机交互 会议 体可视化 体绘制

信息可视化 动态图可视化

动态网络 北京大学 可视分析

可视化 可视化前沿讲习班 可视化应用

图可视化 城市数据 多层次 大规模数据可视化

微博可视化 散点图 数据立方体 数据降维

文本可视化 时变数据可视化 时序数据

时空信息可视化 时空可视化 暑期学校 机器学习

比较可视化 深度学习 深度神经网络 科学可视化

移动数据 聚类 高维数据 高维数据可视化

近期文章

通过协同语义推理与深度学习模型可视交互 (Visual Interaction □with Deep Learning Models □through Collaborative Semantic Inference)

基于射线代理分布的大规模时变体数据探索 (Ray-based Exploration of Large Time-varying

Volume Data Using Per-ray Proxy Distributions)
学术实体的影响花朵 (Influence Flowers of Academic Entities)
流式多维数据可视化的增量降维方法 (An Incremental Dimensionality Reduction Method for Visualizing Streaming Multidimensional Data)
GUIRO: 用户引导的矩阵重排序 (User-Guided Matrix Reordering)

近期评论

Zhang Wentao发表在《第六届中国可视化与可视分析大会 (ChinaVis 2019) Day 2》
by发表在《基于话题探索和嵌入式可视化促进研究思路生成 (Topic-based Exploration and Embedded Visualizations for Research Idea Generation)》
李, 金城发表在《Clustrophile 2: 可视化指导聚类分析》
Xiaoru Yuan发表在《IGAL可视化讲习班-2: 未来工作模式: 使用先进可视化技术提升交流与合作 – 赵健》
Shawn发表在《IGAL可视化讲习班-2: 未来工作模式: 使用先进可视化技术提升交流与合作 – 赵健》

文章归档

- 2019年十二月
- 2019年十一月
- 2019年十月
- 2019年九月
- 2019年八月
- 2019年七月
- 2019年六月
- 2019年五月
- 2019年四月
- 2019年一月
- 2018年十二月
- 2018年十一月
- 2018年十月
- 2018年九月
- 2018年八月
- 2018年七月
- 2018年六月
- 2018年五月
- 2018年四月
- 2018年三月
- 2018年一月
- 2017年十二月

2017年十一月
2017年十月
2017年九月
2017年八月
2017年七月
2017年六月
2017年五月
2017年四月
2017年一月
2016年十二月
2016年十一月
2016年十月
2016年九月
2016年八月
2016年六月
2016年五月
2016年四月
2016年一月
2015年十二月
2015年十一月
2015年十月
2015年九月
2015年八月
2015年七月
2015年六月
2015年五月
2015年四月
2015年二月
2015年一月
2014年十一月
2014年十月
2014年九月
2014年八月
2014年七月
2014年六月
2014年五月
2014年四月
2014年一月
2013年十二月
2013年十一月
2013年十月
2013年九月
2013年八月
2013年七月
2013年四月
2012年六月
2012年五月
2012年三月
2011年十一月
2011年十月
2011年三月
2011年二月

