

# PKU Visualization Blog

北京大学可视化与可视分析博客

## 时变集合模拟数据的有效可视化 (Effective Visualization of Temporal Ensembles)

作者: Fan Hong 日期: 2015年12月9日

随着计算能力的不断提高, 集合模拟 (ensemble) 数据已经被广泛应用于各类科学模拟中。从定义来说, 一个集合模拟数据是一组紧密相关数据集的集合, 每个数据集被称作“成员”, 它们来自于同一次模拟或同一次试验的多次运行结果。通常集合模拟数据的分析难度巨大, 既因为其数据量庞大, 也因为其具有时变、高维、以及多变量的多重复杂性。本文提出了一系列的可视化分析方法来研究这类具有时变特征的集合模拟数据, 尤其是其关注其所包含特征在各成员间的异同及演变过程。

本文所使用的集合模拟数据来自于RHCI模拟。简单来说, RHCI模型用于模拟金原子的撞击过程, 并分析撞击后在高温致密的爆炸中生成的物质与能量。作者希望通过发展一套可视化分析方法来帮助物理学家研究各成员模拟中的粒子所组成的形态及其演变过程。这个数据共包含41个成员, 每个成员包含60个时间步。每个成员包含的粒子数从18万到330万不等。作者称, 尽管文中一些方法是针对三维空间中的粒子数据所设计, 但还是能很容易地扩展到一般性的时变集合模拟数据的。

作为时变分析的基础, 作者首先关注静态集合模拟数据的可视化, 即仅考虑一个时刻 (时间步) 中各成员间特征的比较。在这个应用中, 特征即是由粒子构成的特定形状。为了衡量粒子形态的差异, 作者提出了一个基于八叉树的比较方法。在这个方法中, 整个空间范围被不断地递归地划分成8个子区域, 当有子区域中包含的粒子数少于指定的阈值时停止其划分。这样在这个八叉树中, 每个节点对应一个空间区域, 它包含一定数量的粒子, 作者在这个节点上存储这些粒子的一些统计值用于之后的可视化, 例如粒子的平均位置, 其他属性 (如温度) 的均值与标准差等等。为了计算两棵八叉树的差异, 作者首先定义两棵树在对应结点 (Octant) 上的差异值, 为所包含粒子数的差值除以两者粒子数的较大者。接着, 八叉树中每一层的差异值定义为所有结点差异值的均值。而整个八叉树的差异值则定义为所有层差异值的加权平均值, 这里作者允许使用者通过设置权重来展示八叉树不同层次的细节信息。通过如上的定义, 各成员间的两两差异值就可以计算出来了, 作者利用这个差异值矩阵进行层次聚类 (见后文)。同时, 根据八叉树所存储的信息, 作者提出用符号可视化的方式来展示单个成员, 如图1所示。其中点的大小表示包含在对应结点中的粒子数, 颜色表示用户选定的某属性的均值, 点闪烁的频率表示其标准差 (见论文视频)。

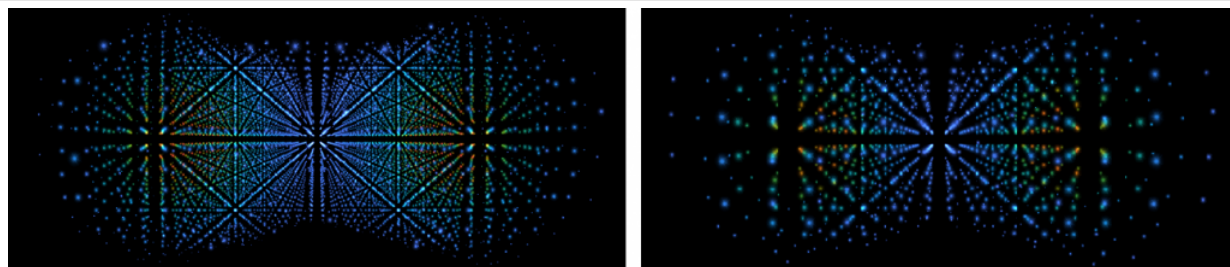
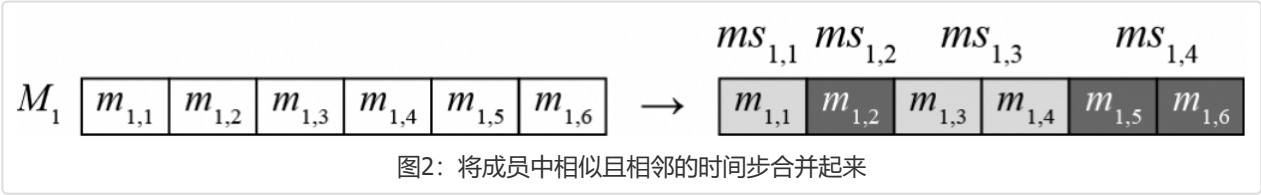


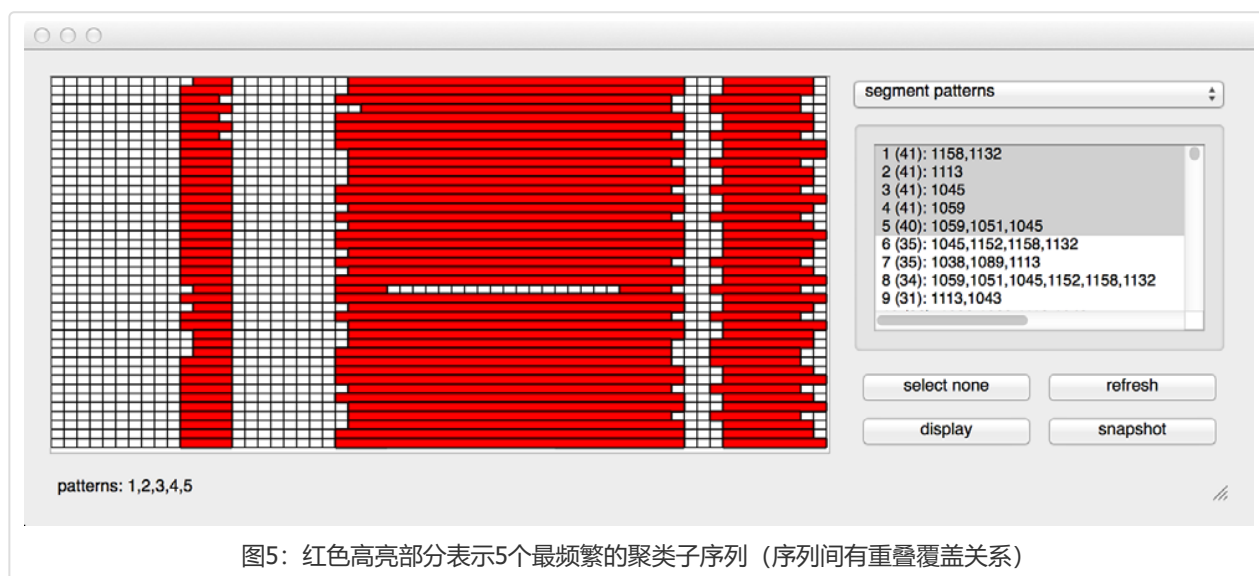
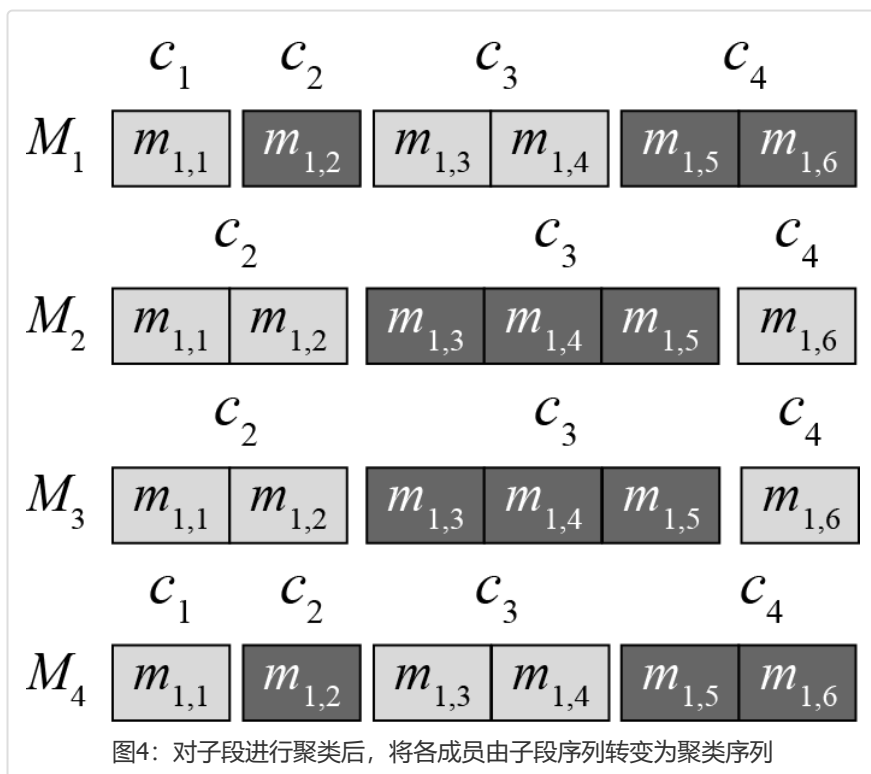
图1: 对于单个时间步粒子形态的可视化。左图包含八叉树的完整结构, 右图仅包含八叉树中第四层的结构。

在静态集合模拟数据可视化的基础上, 作者提出了基于分段的集合模拟数据可视化。当引入时变信息后, 计算任意两成员在任意两时间步上的形态差异是一项十分昂贵的任务。因此, 作者提出首先将各成员中相似且相邻的时间步先合并起来, 如图2所示。对于每个单独的成员, 作者每次选出一个没有被合并的时间步作为当

前的子段，然后向两头扩展，如果它们与当前子段的形态差异小于一定的阈值，就合并进该子段。这样重复进行，直至所有的时间步都有所属子段为止。当然，这里需要计算单个时间步与一个子段（多个时间步）的差异值，本质上仍然是计算两棵八叉树的差异，但对于一个子段来说，其八叉树是其所包含时间步的八叉树的“平均”。



之后，为了分析跨成员间粒子形态的相似性，作者对所有的子段进行了一次聚类，并以此结果将各成员由子段的序列转变成聚类的序列，如图4所示。这样再通过平凡子序列模式挖掘 (CISP) 技术，就能十分方便地分析各成员间粒子形态的相似性与差异性。例如，图5中展示了最频繁出现的5个聚类序列，其中可以很容易地观察到有一个成员缺少其他所有成员都有的一个形态变化阶段。



而另一方面，在对各成员进行分段之后，作者使用动态时间规整算法 (Dynamic Time Warping) 来计算各成员间的差异性，并以此进行聚类操作，来观察各成员在整个事件跨度上的相似性与差异性。

除了基于分段的可视化方法之外，作者所合作的物理学家们也强烈要求实现基于时间步的可视化方法，其目的就是要比较各成员在每个对应时间步上的形态差异。具体实现起来也比较直接，即在每个时间步上应用一次前文提到的静态集合模拟数据的可视化方法即可。图6是对于第50时间步进行凝聚性层次聚类の結果，通过调节参数，作者仅将最后的两个聚类展示了出来，可以看到这两个聚类存在明显的形态差异，即中间部位的断开与连接。当然，在基于分段的可视化方法中所应用的CISP技术，以及对于整个时间尺度的成员进行聚类的方法也可以应用到这里，得到相应的结果。



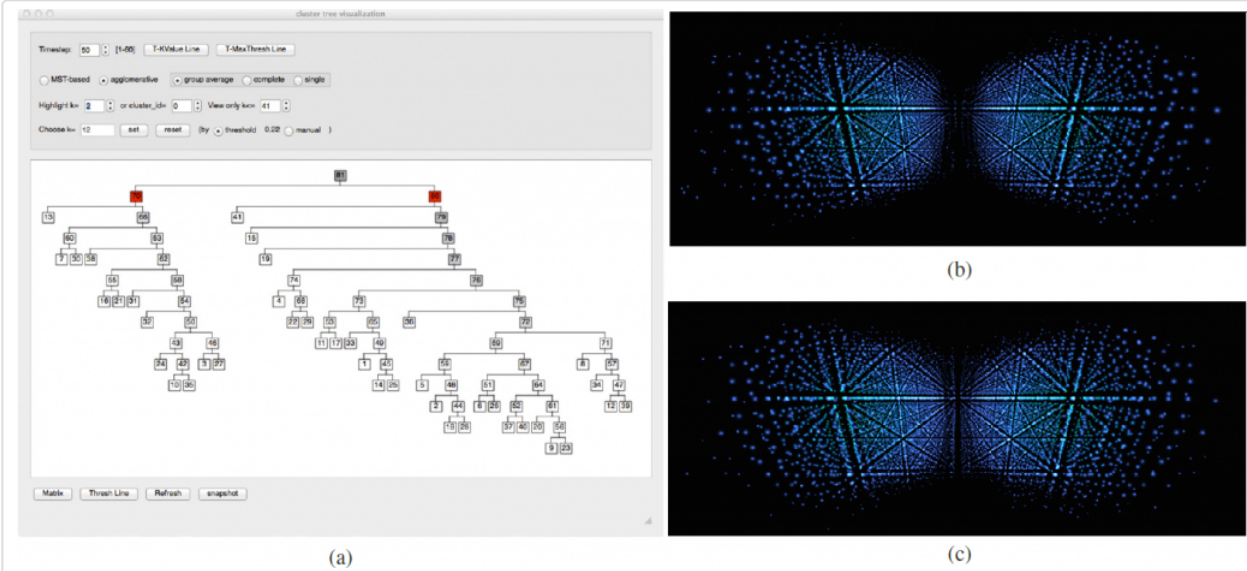


图6: (a)对成员进行凝聚性层次聚类结果; (b,c)选取最后两个聚类进行可视化, 两者在形态上有明显的差别。

总起来看, 作者在这篇文章提出了一系列的方法用于分析时变的集合模拟数据。其中, 基于八叉树的粒子形态比较与基于符号的可视化方法是针对该粒子数据特别设计的, 但仍然具有一定的借鉴意义。而基于分段与基于时间步的两个分析方向, 以及其中所用到的分段方法、频繁模式挖掘、动态时间规整算法、以及层次聚类方法, 尽管都是比较直接的应用, 但都具有很强的普适性。在如今, 已有文献对时变集合模拟数据还关注得不多的情况下, 文中提出的方法对于分析其他时变的集合模拟数据可以发挥出巨大的作用。

[1] Lihua Hao, Christopher G. Healey, Steffen A. Bass: Effective Visualization of Temporal Ensembles. IEEE Trans. Vis. Comput. Graph. 22(1): 787-796 (2016).

论文报告      时变数据可视化, 科学可视化, 集合模拟数据

← 英国帝国理工学院郭毅可教授访问北京大学可视化实验室

科学大数据可视化学术研讨会成功举办 →

评论关闭。

RSS 订阅

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