

PKU Visualization Blog

北京大学可视化与可视分析博客

时变集合模拟数据中的趋势特征可视分析 (Visual Trends Analysis in Time-Varying Ensembles)

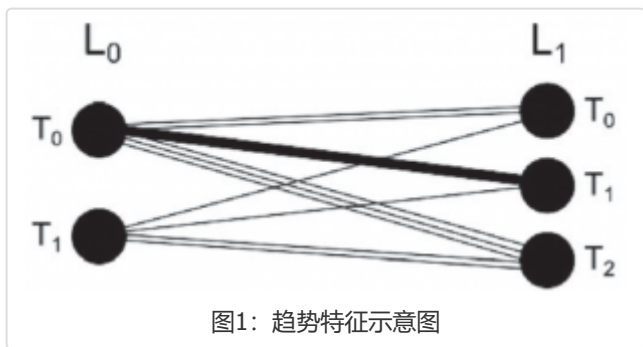
作者: Fan Hong 日期: 2017年7月9日

集合模拟数据是现今科学可视化领域中的重要挑战之一。对于同一个物理现象，使用多个物理模型或者同一模型多组不同参数进行模拟，产生的一组模拟结果就称之为集合模拟数据，每个单独的结果称之为集合成员。对集合模拟数据的研究，一方面可以对各个结果之间的相似性和相异性进行分析，另一方面可以用于进行模型参数的优化。而这个工作主要关注的对应两个问题就是：时变集合模拟数据中趋势特征和异常成员的识别，以及和参数空间的结合探索。

作者首先对时变集合模拟数据中的趋势特征进行了定义。在这个定义中，我们考虑一个空间位置 x 在各个成员上的某个行为。这里，行为的概念可以比较灵活的使用：在标量场中可以直接定义为这个位置所具有的标量值，在流场中可以认为是从这个位置放置粒子在一定时间平流之后的位置，等等。总之，这个“行为”最终可以用看作是一个一维、二维、或者更高维度的值向量。

那么对于任一时刻 t 和任一位置 x ，我们根据各个成员上在该位置上的行为进行聚类操作，就能将所有成员划分成若干个集合。其中，单元素聚类就可以认为是异常成员。这里，为了避免人工设定聚类的个数，作者采用了基于最小生成树的聚类方法，详细信息可以参看原文章。

接着，基于这个位置 x 在不同时刻上各个成员的聚类结果就可以定义“趋势”特征：如果在一定的时间步区间里，每个时间步选择一个聚类，使得相邻时间步的所选聚类至少共享一个集合成员，那么这个聚类的序列就可以称之为一个趋势特征。例如，在图1中，在时刻 L_0 中有两个聚类 T_0 和 T_1 ，在时刻 L_1 有三个聚类 T_0 、 T_1 和 T_2 。对于不同时刻的同一个成员，用边直接连接。我们可以十分容易地看到， L_0 时刻的任一聚类和 L_1 时刻的任一聚类都共享集合成员；那么，他们都可以称之为这个时间区间下的一个趋势特征。当然，实际分析中，所考虑的时间区间通常会大于2。



对于这个趋势特征的意义，可以从多个方面来解读。一方面，它与参数空间紧密对应。一个通常比较合理的假设是，所提取的集合成员的聚类应该也对应参数空间中参数的聚类，并且这些聚类之间应该能比较好的分隔开。但如果，这个假设被打破，那就意味着模拟中可能有一些有趣的事件发生，例如数值计算带来的误差、一些特殊的物理现象等等。找到趋势特征和异常成员将有助于这类分析。另一方面，趋势特征在时间上

有一致性的属性。如果一个趋势特征所包含的成员在所有时间步上一直没有变化，我们称之为“完美一致性”。但如果有成员加入或者退出一个趋势特征，则可能意味着一些重要的知识。

为了帮助使用者从以上两个角度来对趋势特征和参数空间进行分析，作者设计了如图2所示的可视分析系统。这个系统的界面主要包含三个部分：趋势图、参数空间视图、以及集合视图。

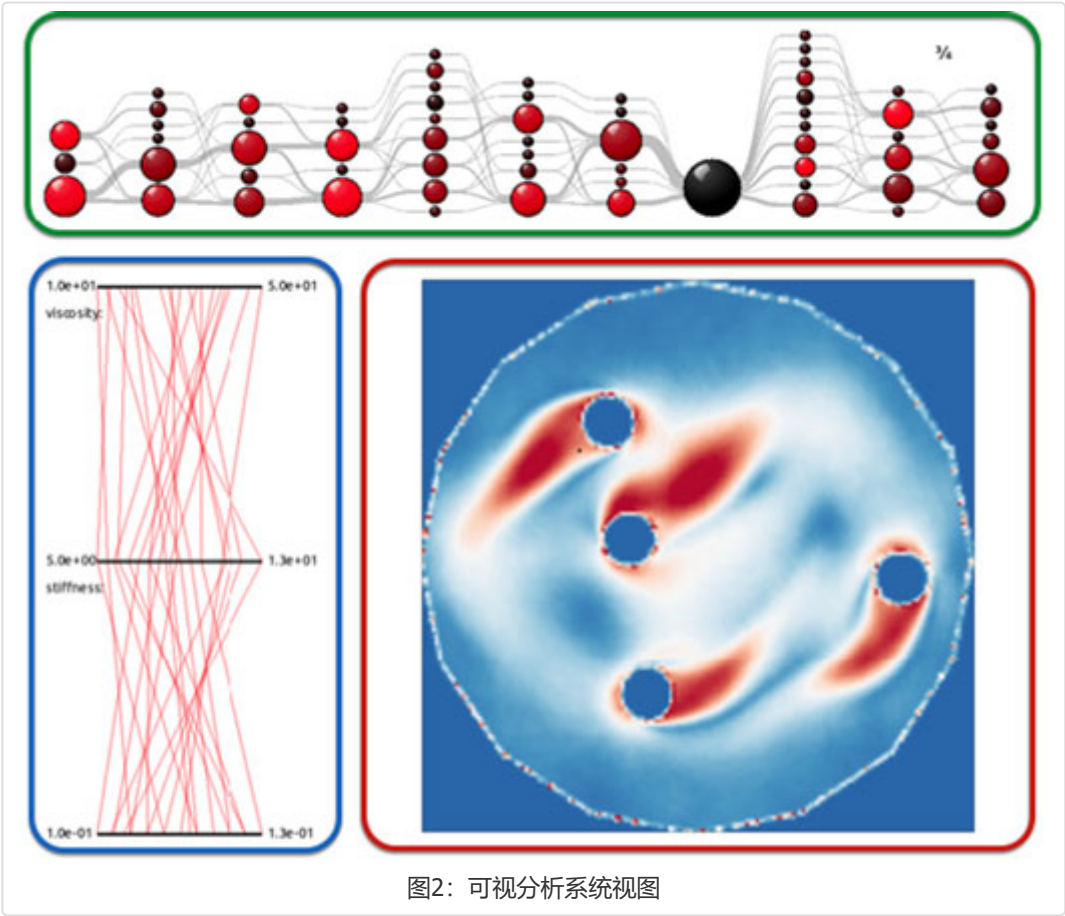


图2：可视分析系统视图

趋势图对某个位置x在各个时间上各成员的聚类结果进行可视化。在图1的基础上，节点大小用来表示对应聚类所包含的成员数量，颜色表示所关心的某个标量值；节点与节点之间的边也通过边捆绑技术进行聚合，用宽度表示所含成员的数量。所得到的结果如图3所示。更进一步地，如果数据本身所含时间步过多，导致趋势图太长，作者还提出多层次的趋势图来缓解这个问题。即在原始的趋势图上，每隔一个时间步抽取一个时间步来组成新的下一层次趋势图，以此类推，如图4所示。

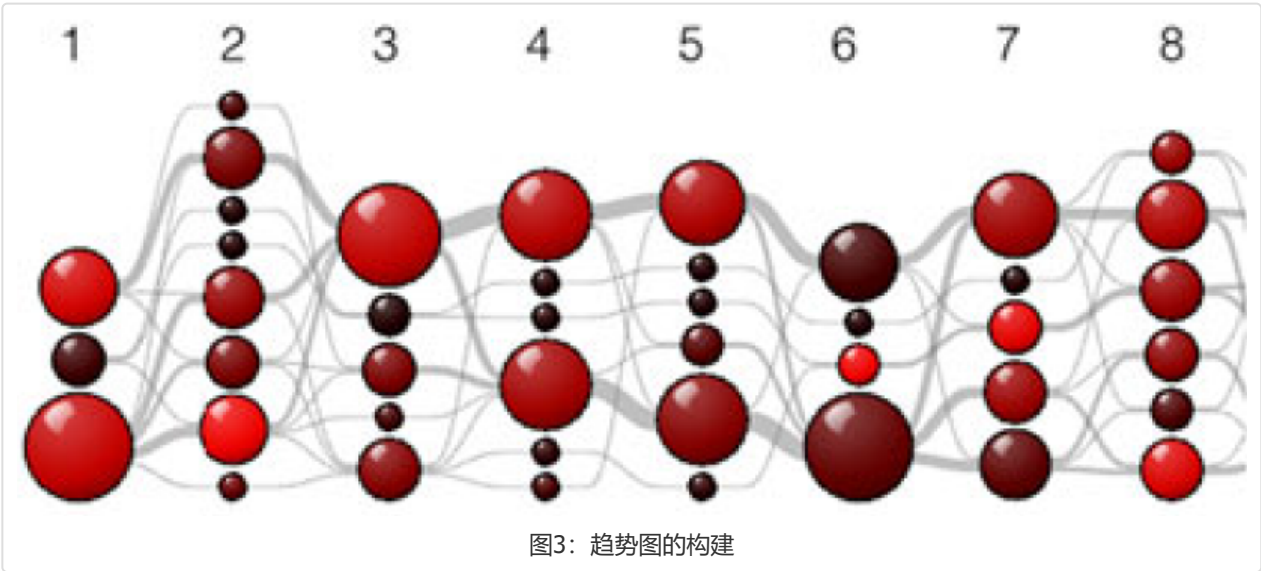
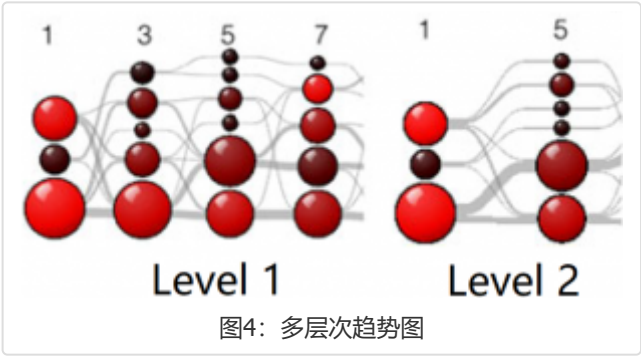
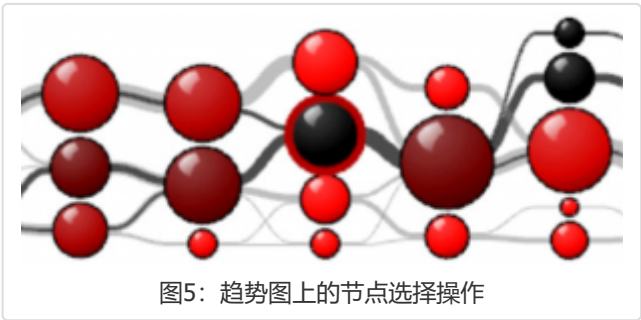


图3：趋势图的构建

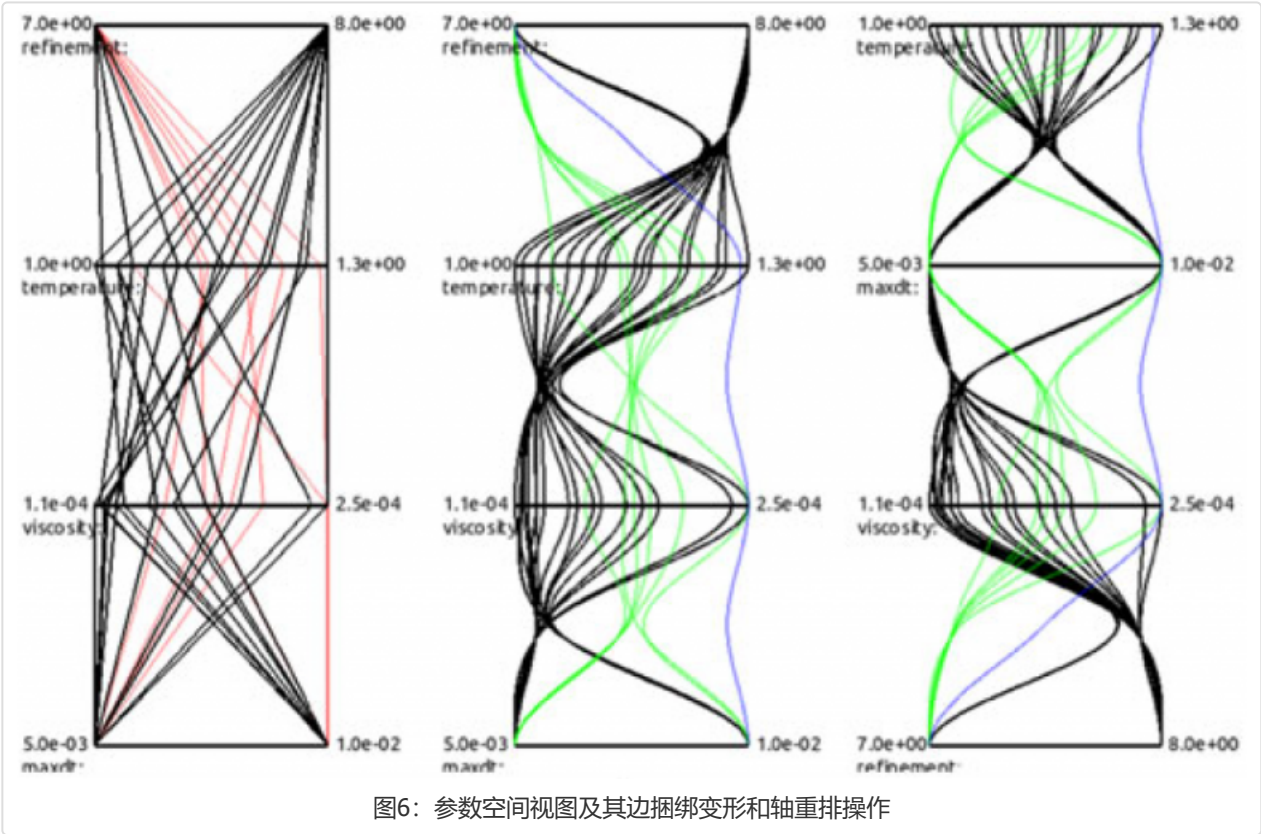


趋势图上的交互主要为单节点选择。对应聚类被选择后，其中所含成员在其他时刻的分布和走向，会用颜色表示出来。例如，图5中第三列第二个节点被选中，用红色外延标记。其中节点在其他时刻的分布以及流向，分别用节点和边的颜色表示，越接近黑色表示在对应的几何里占比越高。



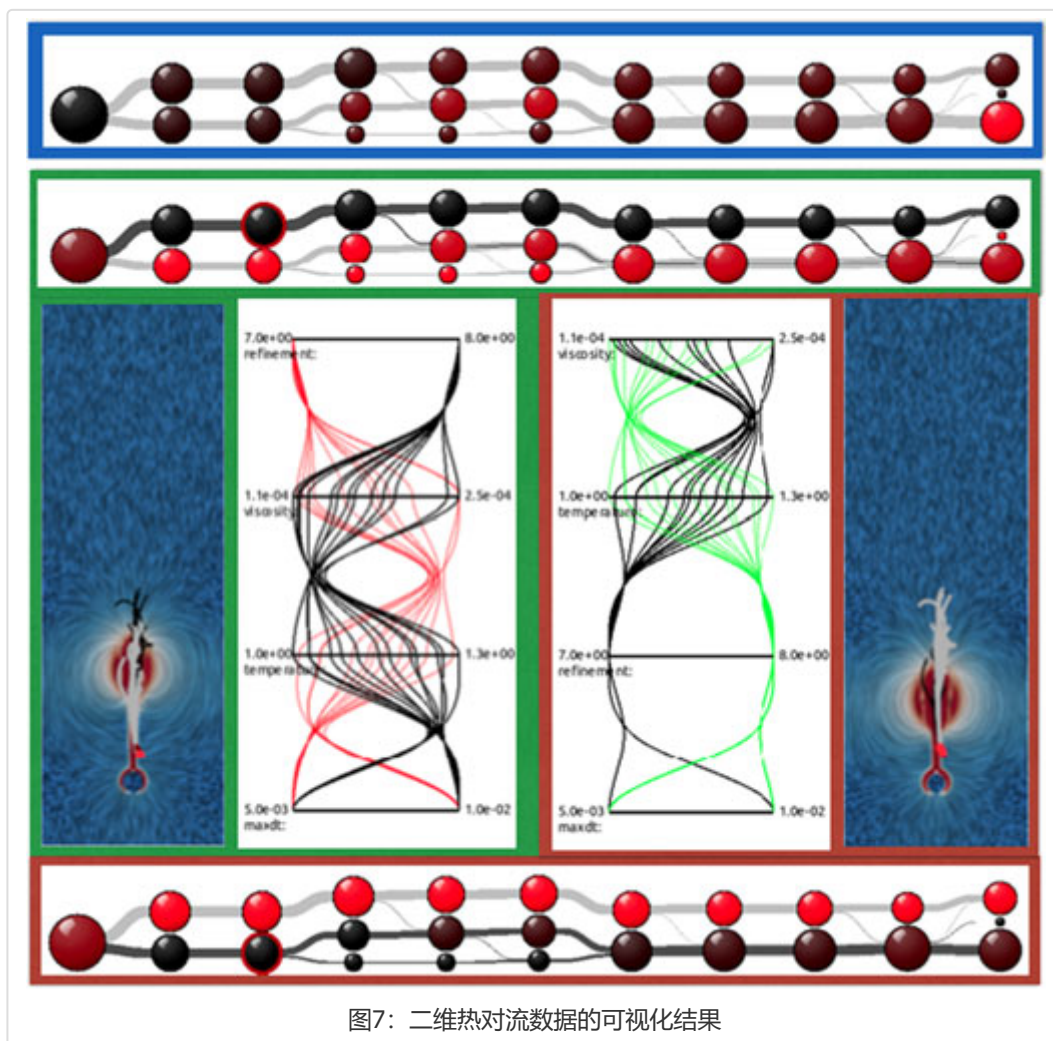
参数空间视图主要就是一个参数空间的平行坐标，如图6所示。每个轴表示一个参数，每条折线对应一个集合成员。这个视图与趋势图可以进行双向的联动选择。对于所有折线，作者也使用边捆绑技术来对边进行聚合，以便于使用者识别参数的线性可分性。

图6



最后，集合视图本质上是一个空间视图，直接对各个空间位置的某个量进行展示，如图7所示。这个量可以是该位置各个成员行为的一个均值，也可以是所含趋势特征的数目。作者还定义了相似性、趋势复杂性等等来刻画趋势的行为。在实际分析中，这个视图实际上是总的入口，使用者从这个视图中选择一个位置，然后开始对应趋势图的分析。

下面简单介绍一个例子：二维热对流数据中流场输运行为的分析。数据包含121个时间步，具有两种网格大小。所含参数包括粘度、温度、模拟时间步的长度、以及网格分辨率。图7展示了作者选择了一个位置（空间视图中的红色点）后进行可视分析的结果。图中最上面的趋势图是总的概况，可以很容易的观察到所有的集合成员最初分成了两部分，其中一部分中间会在分裂一次，之后又汇聚起来形成一个聚类。使用者然后分别选择第三个时间步的两个聚类，查看他们对应的参数空间情况和空间中迹线分布情况，分别用绿色和红色框对应。能从空间视图上很直接地看到两个聚类的区别。而从参数空间视图也能看到两个聚类对应的参数完全由网格分辨率这一参数区分开。



接下来，作者想了解这种完全由网格分辨率来区分的情况是不是在所有位置都存在。作者在参数空间视图中直接选择高分辨率的成员，然后在空间视图中展示各个位置的“相似性”。这里的“相似性”用来衡量对于所选的成员集合，他们有多长的时间属于同一个趋势里，也就是没有分开的情况。从图8中可以看到实际上只有在一部分空间位置（红色表示）中，这些成员会一直在一个趋势中。这个发现也否定了作者一开始的猜想。

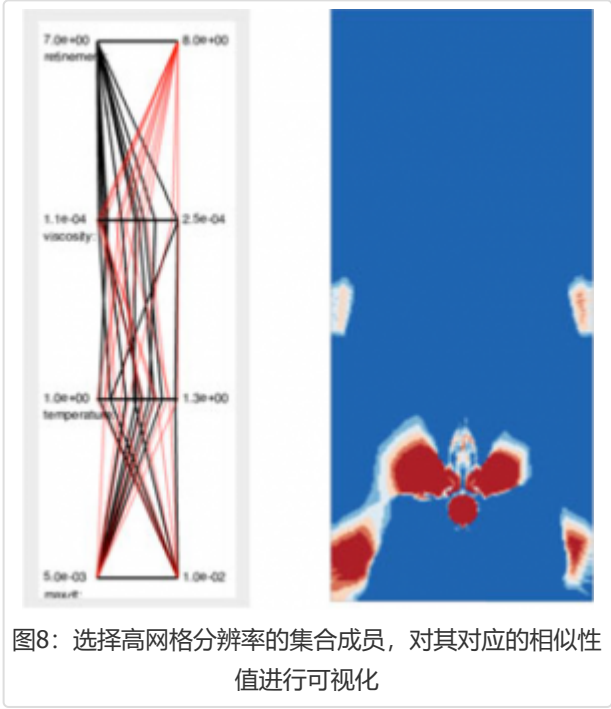


图8：选择高网格分辨率的集合成员，对其对应的相似性值进行可视化

总结起来，本文首先定义了时变集合模拟数据中的“趋势”特征，并设计了一个可视分析系统来进行参数空间的联动分析，能够帮助使用者更容易地了解参数空间。尽管所采用的可视化方法比较简单直接，但在具体的问题上取得了不错的结果。

[1] Harald Obermaier, Kevin Bensema, Kenneth I. Joy. Visual Trends Analysis in Time-Varying Ensembles. IEEE Trans. Vis. Comput. Graph. 22(10): 2331-2342 (2016)

论文报告 时变数据, 聚类, 趋势, 集合模拟数据

← Screenit: 对细胞筛选的可视分析 (Screenit: Visual Analysis of Cellular Screens)

首届中日可视化研讨会7月24日在京举办 →

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