

PKU Visualization Blog

北京大学可视化与可视分析博客

时变数据分析中基于分布的特征提取和追踪 (Distribution driven extraction and tracking of features for time-varying data analysis)

作者: Jiang Zhang 日期: 2016年1月8日

特征的提取和追踪是流场可视化中一种非常重要的技术，可以让科学家们直观地理解模拟数据的特性，从而发现有意义的物理现象。大多数已有的特征追踪技术都会通过设置阈值或查询范围等方式事先定义特征，然后再对具体的特征进行提取和可视化。但是，随着数据变得越来越复杂，科学家们经常只能模糊地定义感兴趣的特征，例如涡旋中心和地震冲击波等。此时包含特征的区域不能通过精确地定义来描述，因此科学家们需要运用新的算法来有效地提取和追踪这类模糊的特征。去年的IEEE SciVis上就有一篇文章提出了基于分布的方法，使用高斯混合模型(GMM)对目标特征区域进行表示[1]。该方法对每个数据块在初始时间步进行GMM建模，然后使用增量学习的机制更新数据块GMM在每个时间步的参数。与此同时，对每个数据块在每个时间步计算两种概率，包括该数据块有潜在特征运动的概率和有指定特征存在的概率。将这两种概率进行线性结合，建立新的特征分类场(feature-aware classification field)，然后就可以进行特征的提取和追踪。下面将详细介绍这个工作。

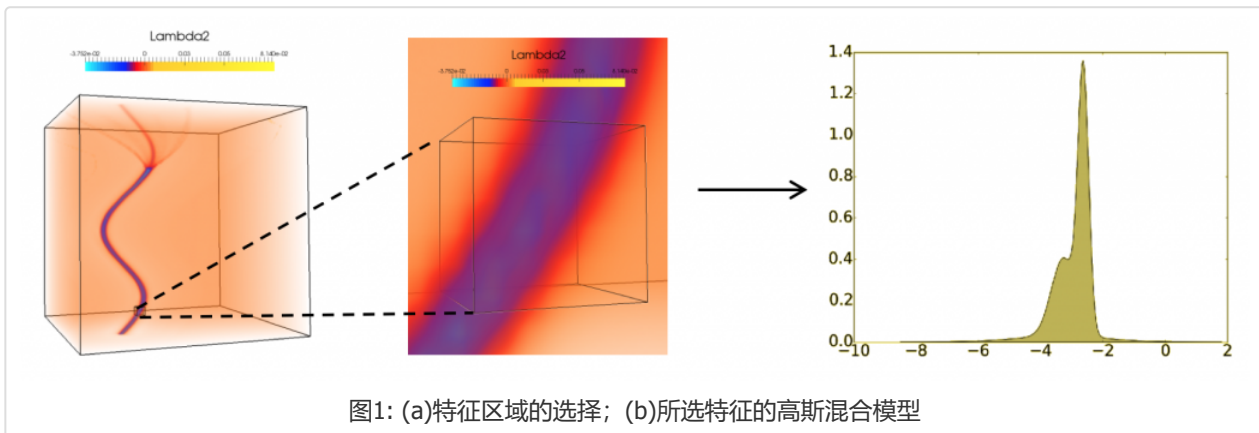
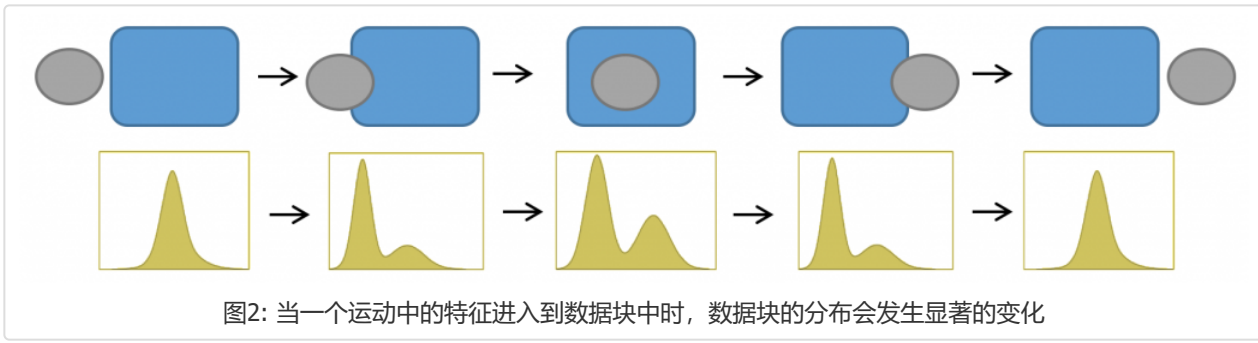


图1: (a)特征区域的选择; (b)所选特征的高斯混合模型

由于没有精确的特征定义，特征选择需要根据科学家们的经验和知识，从初始时间步的原始数据中直接选取感兴趣的区域，然后将该特征区域使用GMM表示，如图1所示。此外，整个时变数据的分布按照块为单位，使用增量的GMM学习方法对每个数据块在每个时间步进行建模。具体来讲，将整个数据域划分为数据块，在初始时间步，将每个数据块使用GMM来表示。在随后的时间步中，不断更新每个数据块的GMM参数。其依据是，检查数据块中每个新数据点是否与其任何现有的高斯分布相匹配。如有匹配，则选择最为匹配的高斯分布并更新其均值和标准差；如没有匹配，则用一个新的高斯分布替代原有高斯分布中最不可能的，新的高斯分布使用该点的数据值作为均值，并且使用高的标准差和低的权重。匹配完成后，更新数据块的每个高斯分布的权重。



在增量学习数据块GMM参数的过程中, 同步计算每个数据块中有潜在特征运动的概率和先前指定的特征存在的概率。前一个概率的评估思想是, 当一个运动中的特征进入到数据块中时, 由于新的数据点的出现会导致其分布发生显著的变化, 例如新的高斯分布会产生, 如图所示2。这类新数据点要么是不与现有的任何权重高于给定阈值的高斯分布相匹配, 要么与新产生的高斯分布匹配。这些新数据点与数据块中所有数据点的数量比值即是数据块中有潜在特征运动的概率。后一种概率评估了每个数据块与指定特征的相似度, 其是通过计算每个数据块的 GMM与指定特征的GMM的距离得到的。两种概率的计算公式如图3(a)和(b)所示。这两种概率每一种都不能单独精确地预测指定特征出现的概率, 因此该工作将两者进行线性结合, 如图3(c)所示, 每个时间步每个数据块都由一个概率值来表示, 从而可以构建一个特征分类的标量场。线性结合的概率对应的结果如图3下方所示, 可以明显看出比单独的概率包含的信息准确全面。

$$(a) Possibility_{foreground,t}(b_{i,t}) = q_{i,t} / n_{i,t}$$

$$(b) Possibility_{similarity,t}(b_{i,t}) = 1 - \psi_{norm}(b_{i,t}, f_t)$$

$$(c) Possibility_{feature}(b_i) = \gamma * Possibility_{similarity}(b_i) + (1 - \gamma) * Possibility_{foreground}(b_i)$$

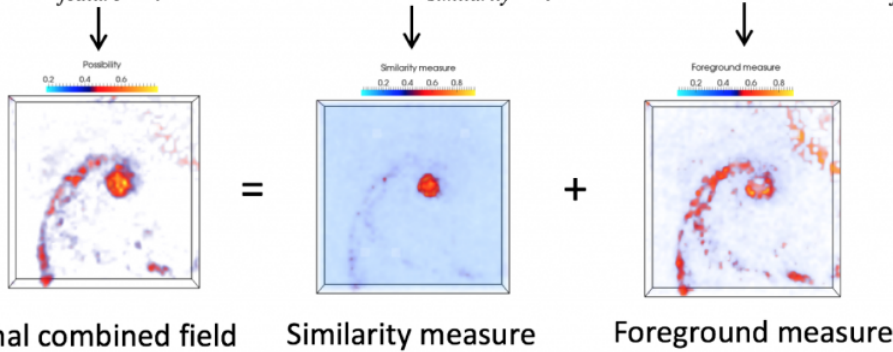


图3: 每个数据块中有潜在特征运动的概率(a)和有先前指定的特征存在的概率(b); 两者线性结合产生新的分类概率(c)

有了特征分类场的计算, 现在可以对指定的特征进行提取和追踪了, 整个过程如图4所示。首先在初始时间步指定目标特征, 然后对之后的每一个时间步, 生成特征分类场, 使用给定阈值对非类场过滤出符合条件的区域并对结果使用区域增长算法提取所有的连通特征区域, 再计算目标特征和所有检测到的区域的距离, 从而得到最为匹配的区域并将其作为新的目标特征, 并在下一个时间步的循环中使用。通过这一过程, 可以追踪到目标特征随时间的变化。

Algorithm 2 Tracking In Feature-Aware Classification Field

```

1: Input:  $GMM(b_i), GMM(feature) : \forall i \in 1, 2, \dots, n$ 
2: Initialize  $f_{target} := GMM(feature)$ 
3: for all  $t$  in  $T$  do
4:   Generate Feature-Aware Classification Field( $GMM(feature)$ ,  $t$ ) (Algorithm 1)
5:   Thresholding ( $\geq poss_{th}$ ) on the Classification field.
6:   Apply connected component algorithm on the thresholded results.
7:   Compute distance between centers of target feature and all the detected regions
   from the current time step.
8:   Find the best match  $l$  with the minimum distance to the target feature  $f_{target}$ .
9:   Set  $f_{target} := l$ 
10: end for

```

图4: 利用特征分类场进行特征提取和追踪的算法综述

文章使用了一些案例来证明方法的有效性。下图展示了Isabel飓风数据的例子。图5(a)在风眼处选择了一个涡旋特征区域，对应的GMM如图5(b)所示。该特征的提取和追踪如图6(a)-(c)所示，涡旋持续向右上角移动，这一观察也与实际情况相符合（图6(d)）。此外，图7展示了与精确特征定义方法进行了比较，可以看出特征的提取结果非常相似。

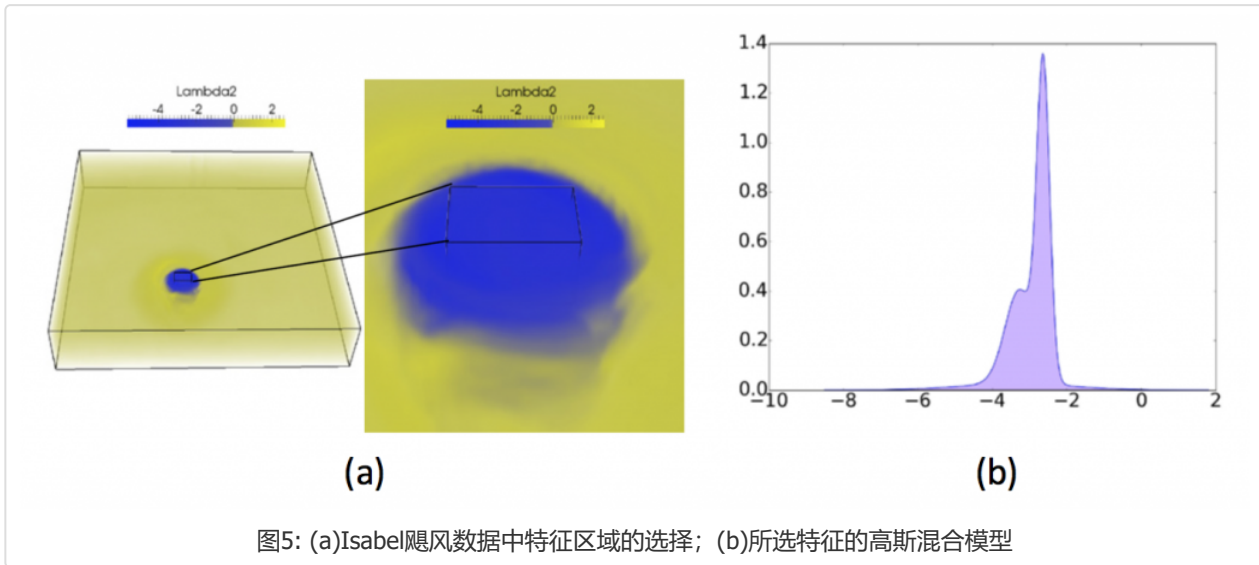


图5: (a)Isabel飓风数据中特征区域的选择; (b)所选特征的高斯混合模型

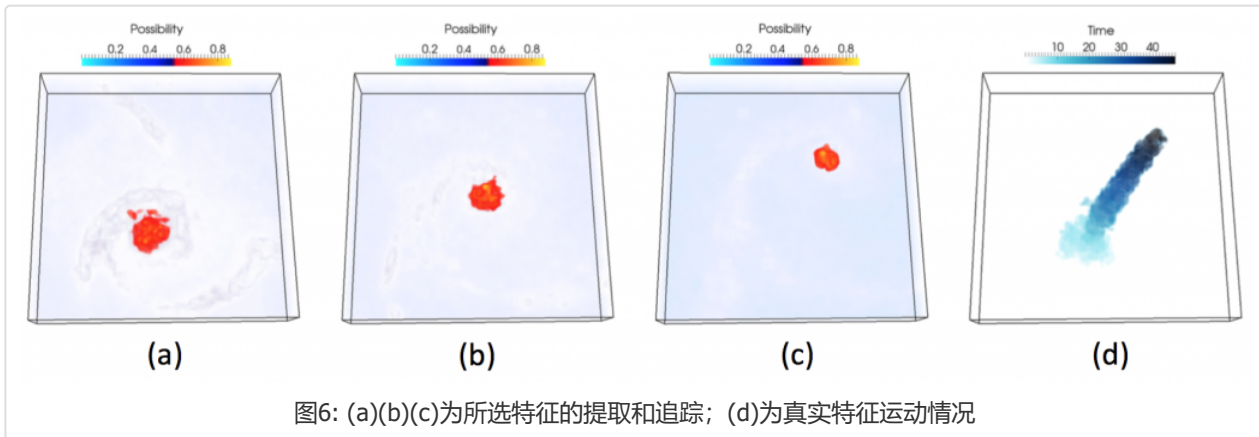
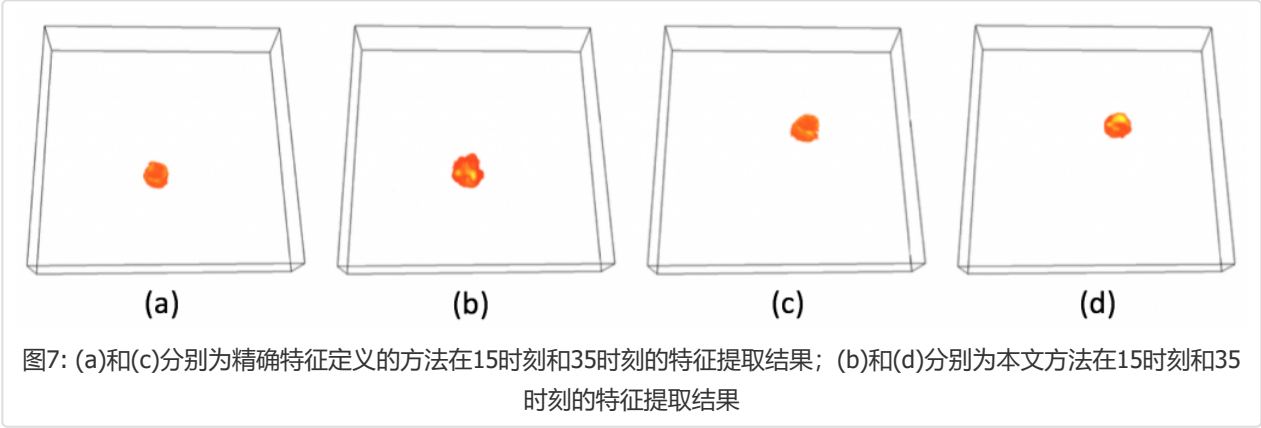


图6: (a)(b)(c)为所选特征的提取和追踪; (d)为真实特征运动情况



基于分布的特征提取和追踪方法使用了增量学习的方法，与离线对每一时间步建模的方法相比大大减少了计算量。并且当目标特征改变时，算法不需要访问原始数据，仅仅需要在初始时间步对特征的GMM进行评估，然后利用已经计算出的GMM就可以完成新的特征提取和追踪。另外，GMM的存储复杂度也远比原始数据开销小得多。唯一不方便的是目标特征的选择需要领域知识，对于不熟悉数据的人来说比较麻烦。该方法也可以比较容易地用于流数据(streaming data)或着原位分析计算框架。

Reference

[1] Soumya Dutta and Han-Wei Shen. Distribution Driven Extraction and Tracking of Features for Time-varying Data Analysis. IEEE Transactions on Visualization and Computer Graphics, 22(1): 837-846, 2016.

论文报告 增量学习, 时变数据分析, 特征提取和追踪, 高斯混合模型

← 多尺度、跨地域的多变量数据可视化 (Visualizing Multiple Variables Across Scale and Geography)

CAST: 基于情境感知的三维点云目标高效拾取技术(CAST: Effective and Efficient User Interaction for Context-Aware Selection in 3D Particle Clouds) →

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
- 新闻
- 未分类
- 活动