点集的贝蒂数

Hao Wang Ratidar Technologies
LLC* 通讯作者：haow85@live.com

摘要

拓扑学是许多工业应用的基础，应用范围涵盖从计算机辅助设计（CAD）到仿真分析。计算拓扑学主要关注网格等结构化数据；然而，点集等非结构化数据集对拓扑学研究者来说仍是一片未开垦的领域。基于点的拓扑学的重要性怎么强调都不为过，尤其是在逆向工程、几何建模和算法分析领域。本文提出了一种计算点集数据贝蒂数的新方法，并通过真实世界示例展示了其用途。据我们所知，我们的工作开创了计算拓扑学领域的先河，是此类研究中的首创。

关键词：贝蒂数、点集、算法分析、拓扑学、计算拓扑学

1. 引言

拓扑学是过去数百年来最成功的数学领域之一。深入理解拓扑学有助于人类探索高维现象和先进物理学。自信息技术革命开始以来，人们便利用计算机在机器上实现拓扑学思想。计算拓扑学广泛应用于仿真分析、计算机辅助设计等领域。然而，与人工智能等其他技术相比，由于理论复杂，这项技术只有一小部分专家掌握。大多数计算拓扑学算法都局限于结构化数据集，例如基于网格的标量场或向量场。点云等非结构化数据集则大多被研究界忽视。

点云之所以重要，是因为它们是由三维扫描仪和雷达采集的数据的原始形式。自动驾驶汽车、建筑信息模型、三维打印以及众多其他工业应用，都高度依赖对点云的有效处理与建模。20世纪90年代，一门名为基于点的图形学的新兴研究领域应运而生，此后十余年蓬勃发展。然而，基于点的图形学基本忽视了拓扑学这一重要领域，转而强调非拓扑建模、渲染等其他方面。据我们所知，关于贝蒂数计算等重要的基于点的拓扑学课题，尚无具有重大影响的研究论文，这对整个科学界而言实属憾事。计算拓扑学的最新发展主要聚焦于研究结构化数据集上的拓扑量和结构，例如标量场上的莫尔斯-斯梅尔复形，以及向量场上的拓扑概念。

构建基于点的拓扑理论，主要难点在于如何量化数学量。点的存在通常与噪声的存在相关，因此会给问题的表述引入不确定性。尽管面临诸多问题，本文仍在这一领域迈出了第一步，提出一种计算点云贝蒂数的新方法。在概述我们的数学表述之后，我们将通过真实世界的实例展示这一新思路的应用。

2. 相关工作

计算拓扑学是一个历经数十年发展的领域。Pascucci [1][2][3] 和 Edelsbrunner [4][5]等研究人员发展了计算莫尔斯理论以及向量场拓扑分析方法，这些方法已广泛应用于制造业和仿真分析领域。尽管贝蒂数等拓扑学基本概念

以及莫尔斯-斯梅尔复形已广为人知，但研究界基本忽视了点数据方面的相关工作。

点数据算法主要是计算机图形学 [6][7] 和计算机视觉 [8][9][10]领域的研究成果。该领域的开创性工作包括移动最小二乘法 [11][12][13], 等。点数据算法广泛应用于逆向工程和自动驾驶汽车行业。

在金融工程领域，极值理论 [14][15] 是一个旨在捕捉金融行业极端事件的子领域。由于幂律效应几乎存在于现实世界的各个方面，这一思想也很容易应用于机器学习。本文在实验中使用了矩估计量 [16] , 、QQ估计量 [17] 和Peng 估计器 [18]。

3. 贝蒂数

对贝蒂数最直观的解释是流形中的孔洞数量。例如，球面的贝蒂数为 1。我们的方法首先计算每个候选孔洞内切圆的半径，然后计算该半径列表的马太效应程度。我们将计算得到的尾指数估计量作为贝蒂数的统计近似。

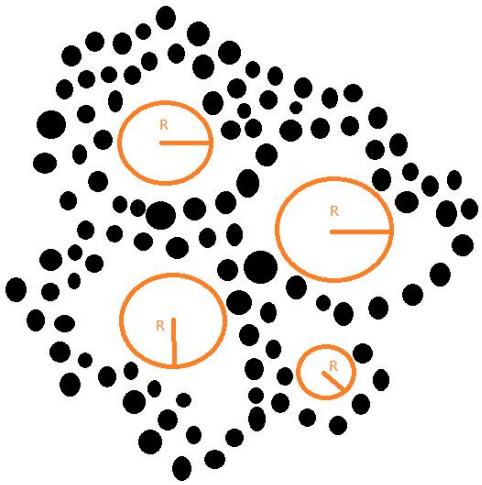

图 1 贝蒂数为 4 的点集

图 1 展示了一个贝蒂数为 4 的点集。我们计算数据集中每一对点之间的距离，取其一半作为潜在孔洞半径的候选值。如图 1 所示，4 个孔洞的半径都包含在候选池中。我们按数值对这些半距离进行排序，然后借鉴金融工程中的极值理论，计算贝蒂数的统计估计值。孔洞的数量相对于点云中的点数较少，而且孔洞的半径远大于其他点对之间的距离。这一现象与幂律效应相符，因此我们选择尾指数估计量来估计贝蒂数。极值理论量化幂律分布的幂指数，是衡量幂律分布偏度的准确指标。

在下一节中，我们将通过几个经典的流形示例阐释计算贝蒂数的思路，然后展示这一思路在真实世界示例中的应用。

4. 流形示例

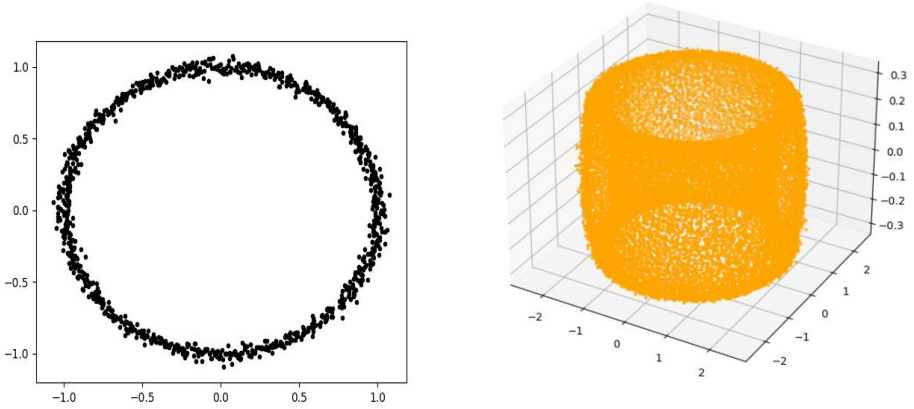

图 2 展示贝蒂数在点云中的概念的两个流形示例——二维圆和三维环面。

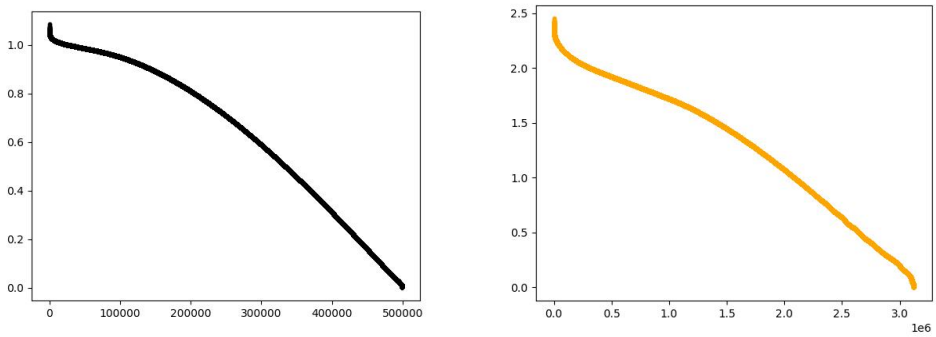

图 3 成对距离一半取值的分布。

图 2 展示了两个基于点的流形示例——带噪声的二维圆和带噪声的三维环面；图 3 展示了这两个流形的排序后的半径列表的分布。我们计算了这两条曲线的极值估计量，并得到以下结果：

	矩估计量	QQ估量	Peng 估计器
圆	-0.62650	0.00140	1.00146
环面	-0.28842	0.00195	1.00195

表1 圆和环面的极值估计量

从表1可以看出，三维环面的贝蒂数估计值大于二维圆。在我们的示例中，三维环面的贝蒂数真实值应为 2，而二维圆的贝蒂数真实值为 1。尽管我们的贝蒂数极值估计量不够精确，但它可作为一种比较指标。

5. 真实世界数据集



图4 点集的真实世界示例

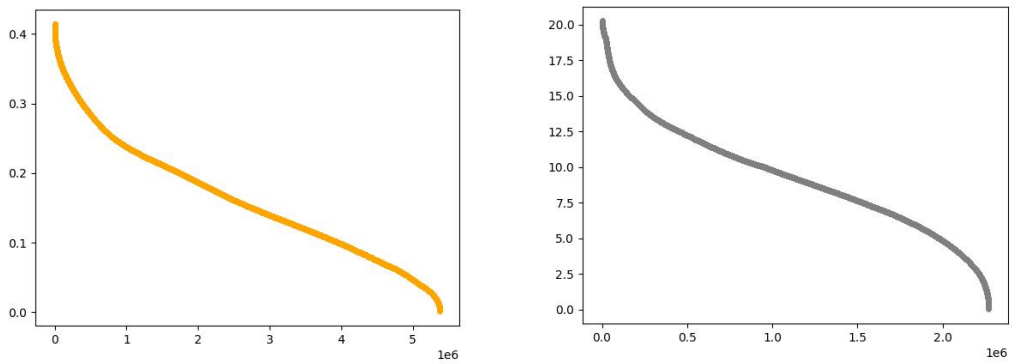

图5 点集的真实世界示例

我们在图4所示的真实世界数据集上测试了这一新颖想法，并在图5中绘制了成对距离的一半。我们在表2中展示了极值估计量的结果：

	矩估计量	QQ估计量	Peng 估计器
Cactus	-0.34205	0.00124	1.00238
Tool	-0.77971	0.00054	1.00083

表2 真实世界示例的贝蒂数

实验结果表明，Tool 的贝蒂数小于 Cactus，这与我们对点云的观察一致。我们已经证明，我们的方法不仅适用于合成示例，也适用于真实世界数据集。

6. 讨论

从实验中我们发现，真实值的贝蒂数越大，贝蒂数估计器也越大。贝蒂数估计器本身并不能精确反映真实值，但它是一种用于比较、排序和排名任务的度量指标。

极值理论中有许多不同的尾指数估计量，但并非所有估计量都能在我们的方法中稳定使用。我们建议读者尝试五种以上不同的估计量，并以多数投票的结果作为最终结果。除本文提及的估计量外，著名的估计量还包括 Pickands 估计器、Hill 估计器和改进 Hill 估计器。事实上，本文中的实验结果已经是从候选估计量集合中选出的多数票结果。

我们的算法也可以应用于算法分析。例如，我们可以对推荐系统的 MAE / RMSE 曲线应用 Takens 嵌入 [19]，将结果映射到二维或三维空间中，再针对点云计算贝蒂数估计器，从而了解算法输出的拓扑复杂性。

我们注意到，pr　　噪声的存在与数据集的离散性一样，是一个主要难题。然而，我们的
该方法能够抵　　御数据集中噪声的存在，如前两节所示。

7. 结论

本文探讨了一种针对点云计算贝蒂数估计器的新思路。我们借鉴金融工程领域的思路，以便计算我们的度量指标，并通过实验表明，我们提出的新方法是有效的。

未来，我们希望探索如何计算点数据上的其他拓扑量和结构，例如莫尔斯-斯梅尔复形。我们还希望研究该问题中的不确定性问题。

参考文献

[1] A. Gyulassy、V. Natarajan、M. Duchaineau 等，《拓扑清洁距离场》，IEEE TVCG，2007年。[2] A. Gyulassy、V. Natarajan、V. Pascucci 等，《三维标量函数的莫尔斯-斯梅尔复形高效计算》，IEEE TVCG，2007年。[3] G. Weber、P.T.Bremer 等，《拓扑景观：科学数据的地形隐喻》，IEEE TVCG，2007年。

[4] D. Cohen-Steiner、H. Edelsbrunner、H. Harer，《持久图的稳定性》，《离散与计算几何》，2007年[5] H. Edelsbrunner、D. Letscher 等，《拓扑持久性与简化》，《离散与计算几何》，2002年[6] K. Yang、X. Chen，《通过点云联合分割进行长方体形状抽象的无监督学习》，SIGGRAPH，2022年[7] G. Metzer、R. Hanocka 等，《通过偶极子传播确定点云方向》，SIGGRAPH，2021年[8] J. Tang、Z. Gong 等，《LAKe-Net：通过定位对齐关键点实现拓扑感知点云补全》，CVPR，2022年[9] Y. Zeng、Y. Qian 等，《IDEA-Net：通过深度嵌入对齐实现动态三维点云插值》，CVPR，2022年[10]Y. Tang、Y. Qian 等，《WarpingGAN：变形多个均匀先验以生成对抗式三维点云》，CVPR，2022年[11] D. Levin，《移动最小二乘法的逼近能力》，《计算数学》，1998年[12] R. Kolluri，《可证明优良的移动最小二乘法》，《ACM算法汇刊》，2008年[13] J.Manson、S. Schaefer，《移动最小二乘坐标》，《计算机图形学论坛》，2010年[14]J. Segers，《极值指数的广义 Pickands 估计量》，《统计规划与推断杂志》，2002年[15]F. Caeiro、I. Gomes 等，《直接缩减经典 Hill 估计量的偏差》，《Revstat—统计学期刊》，2005年[16] M. I. Gomes，《尾指数的广义刀切矩估计量》，Notas e Comunicações CEAUL，1999年[17] M. Kratz、S. Resnick 等，《QQ估计量与重尾》，《统计与随机模型通讯》，1996年[18] V. Paulauskas、M. Vai，《再论尾指数估计量的比较》，《统计学》，2011年[19] H. Wang，《推荐系统的有效可视化与分析》，IFEEA，2022年