

快速聚类算法的性能分析

曹楠 2017310773

January 13, 2018

1 对于论文的中的测试结果进行复现

论文中给出了相当多的实验结果，我们选择了有代表性的 Flame 和 Path-based 数据集进行测试 Flame 数据集由 240 个二维点组成，共有两类，实验结

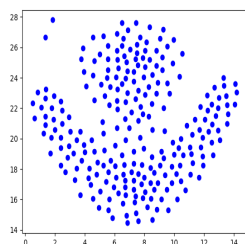


Figure 1: Flame 数据集

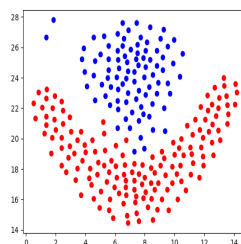


Figure 2: Flame 数据集实验结果

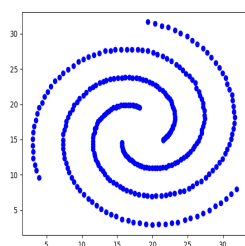


Figure 3: Pathbased 数据集

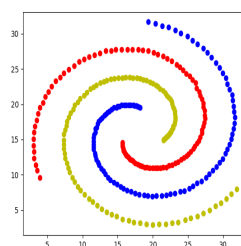


Figure 4: Pathbased 数据集实验结果

果中，240 个点正确分对了 237 个，正确率 98.8%。

实验结果表明，快速聚类算法对 Flame 算法十分有效。

Pathbased 由 300 个二维点构成，共有三类，实验结果中，所有分类都是正确的，正确率 100%。

2 与主流算法进行对比

共在六个数据集进行比较，每个数据集有 1500 个二维点。对比算法有 Mini-Bath KMeans、Affinity Propagation、Mean Shift、Spectral Clustering、Ward、Agglomerative Clustering、DBSCAN、Birch、Gaussian Mixture。

从测试集 1 和测试集 2 中我们可以看到，快速聚类算法依然不能很好的捕捉特殊几何图形的结构，比如圆形、弧线形，

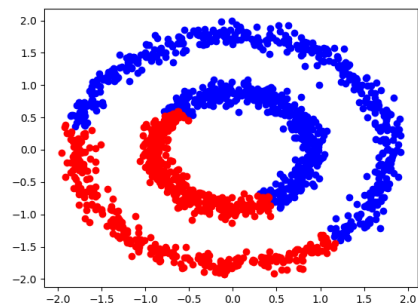


Figure 5: 测试集 1 数据集

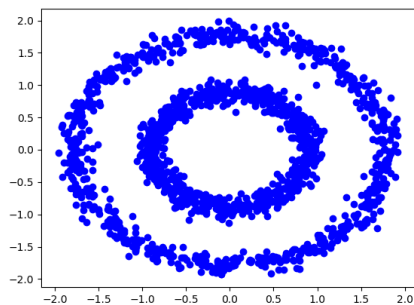


Figure 6: 测试集 1 数据集实验结果

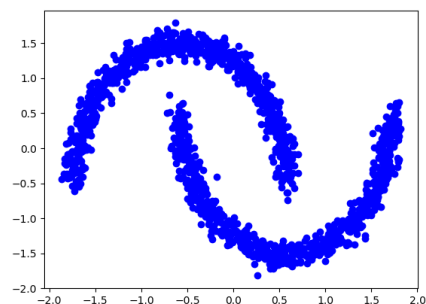


Figure 7: 测试集 2 数据集

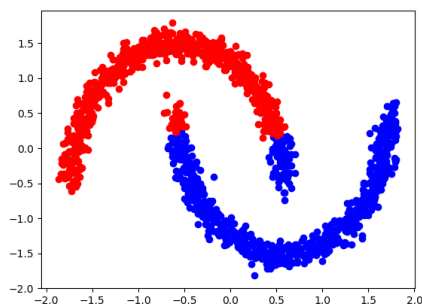


Figure 8: 测试集 2 数据集实验结果

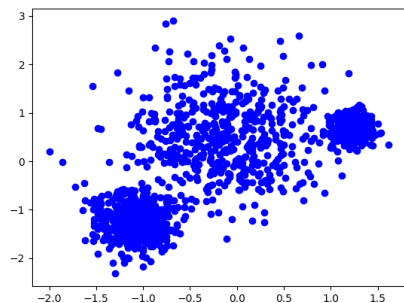


Figure 9: 测试集 3 数据集

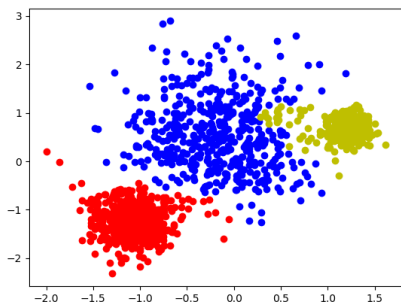


Figure 10: 测试集 3 数据集实验结果

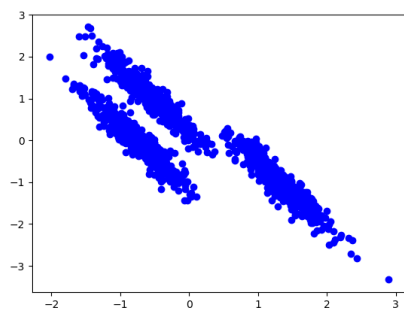


Figure 11: 测试集 4 数据集

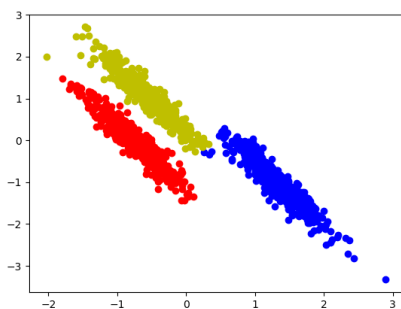


Figure 12: 测试集 4 数据集实验结果

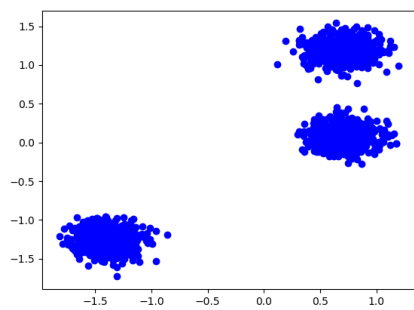


Figure 13: 测试集 5 数据集

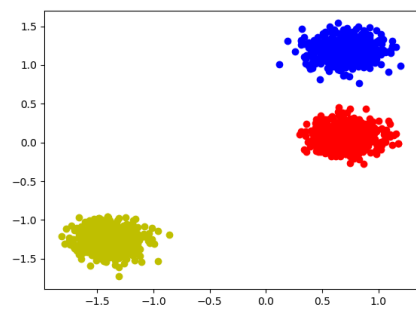


Figure 14: 测试集 5 数据集实验结果

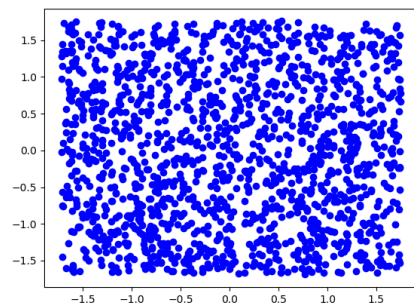


Figure 15: 测试集 6 数据集

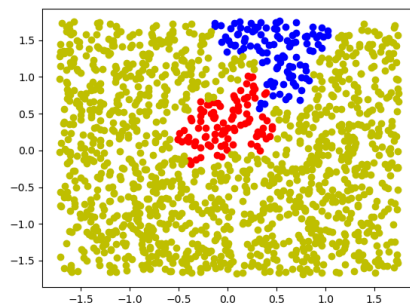


Figure 16: 测试集 6 数据集实验结果

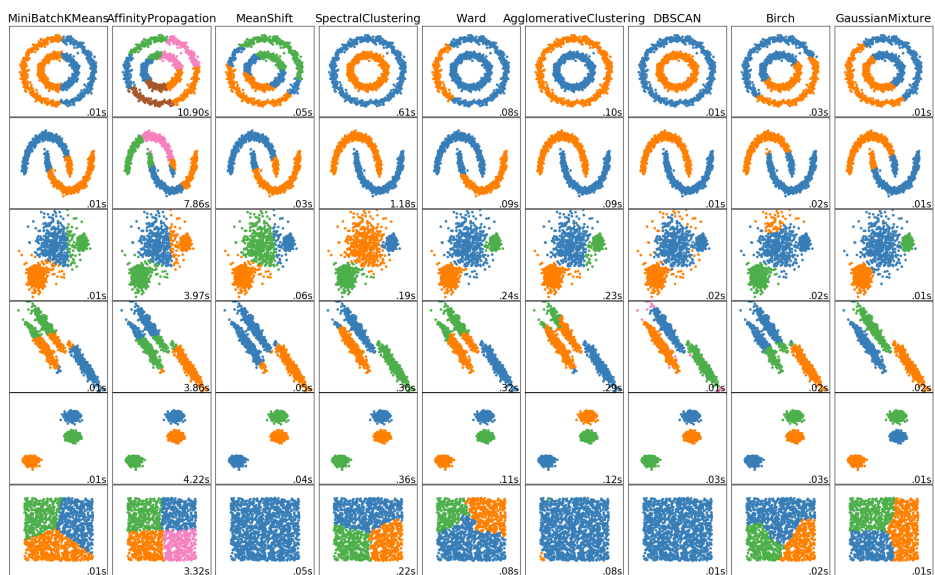


Figure 17: 其他算法对比实验结果