



联合低秩表示和稀疏约束的双层多视角子空间聚类

张福康¹ 尹宏伟¹ 成新民¹ 杜文辉² 徐黄镇²

¹湖州师范学院 信息工程学院 湖州 313000 ²湖州市特种设备检测研究院 湖州 313000

摘要

多视图聚类是利用多个不同描述方式的数据集将同一类相似的数据尽可能聚集到一起。其中多视图子空间聚类是处理高维数据的一种聚类方法。目前的方法是通过分别在每个视图上构造邻接矩阵来解决多视图子空间聚类问题。没有考虑到低秩表示和稀疏约束的结合在构造邻接矩阵中的重要性。为了解决这个问题，本文提出联合低秩表示和稀疏约束的双层多视角子空间聚类方法，使其更全面的描述数据本身,从而实现更有效的聚类。对于每个视图相关的低秩表示和稀疏性约束优化问题，我们使用ADMM方法来解决。最后将在多个数据集上进行实验，发现其聚类性能比现有的多视角子空间聚类算法好。证明低秩表示和稀疏约束的结合可以提高聚类的准确性。

引言

由于数据来源的多样性，同一个数据可能有不同的观察视角。通常，从一个视角得到的数据可以推测出某一方面有用的信息。利用多视角学习可以从这些数据中挖掘出这些隐藏信息。在多视角学习中，多视角子空间聚类利用隐藏在多个视图数据中的互补信息和一致信息来提高聚类性能。本文提出了联合低秩表示和稀疏约束的双层多视角子空间聚类（Double-layer multi-view subspace clustering with joint low-rank representation and sparsity constraint, DMVSC）算法，结构如图1所示。具体来说，我们首先使用矩阵分解提取跨多个视图和多个层的潜在因素，使得这些潜在因素可以有效地捕获数据的结构。并且利用双层结构的特征，每层各自施加低秩表达和稀疏约束，有效捕捉数据的全局和局部结构，提高算法性能。

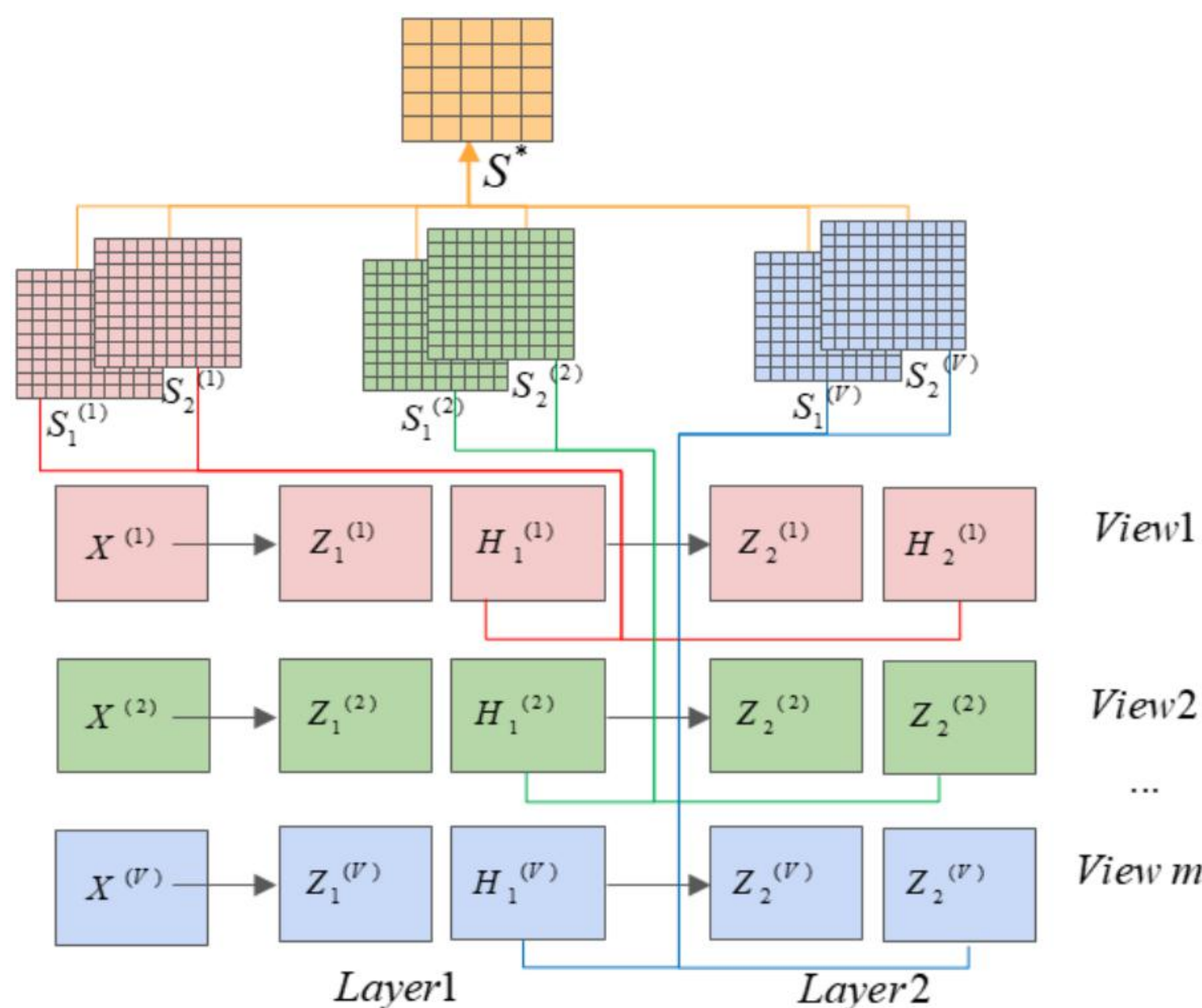


图1 联合低秩表示和稀疏约束的双层多视角子空间聚类结构

注：双层模型的每一层都包含基本矩阵Z和系数矩阵H。Layer1为第一层，用于 $H_1^{(v)}$ 学习低秩子空间 $s_1^{(v)}$ 。Layer2为第二层，用于 $H_2^{(v)}$ 学习稀疏子空间 $s_2^{(v)}$ 。最后，结合的 $s_1^{(v)}$ 的信息获得共生子空间 s^* 。

收敛性分析

为了证明收敛性，图3展示了本文算法在MSRC_v1数据集上的收敛情况。我们可以观察到曲线大约在10个迭代数内达到收敛。证明了我们的算法在实践中可以快速，稳定地收敛。

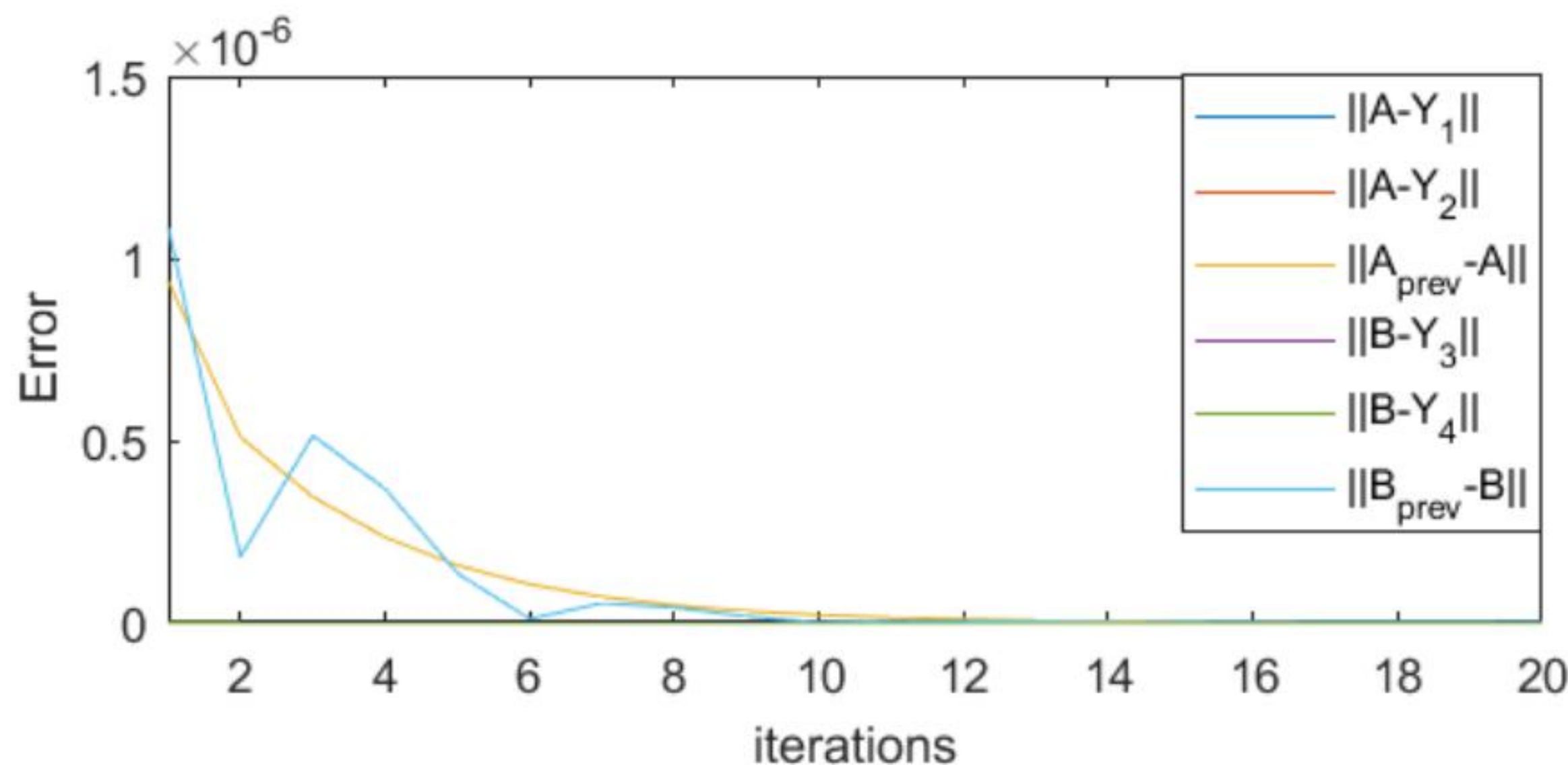


图3 MSRC_v1数据集的收敛曲线

总结

为了解决多视角子空间聚类直接合并会造成多视角数据大量信息的缺失问题。本文提出DMVSC算法，利用矩阵分解产生双层结构将低秩表示和稀疏约束结合在一起，以实现更有效的聚类。最后在实验部分也证明了算法的有效性。DMVSC算法首先在每个视图上执行矩阵分解。之后进一步增强系数矩阵聚类结构:第一层学习低秩表示，第二层学习稀疏约束。优化过程中则使用ADMM方法对目标函数进行求解。最后利用基于质心的正则化方案学习各视图共享的潜在一致数据结构。

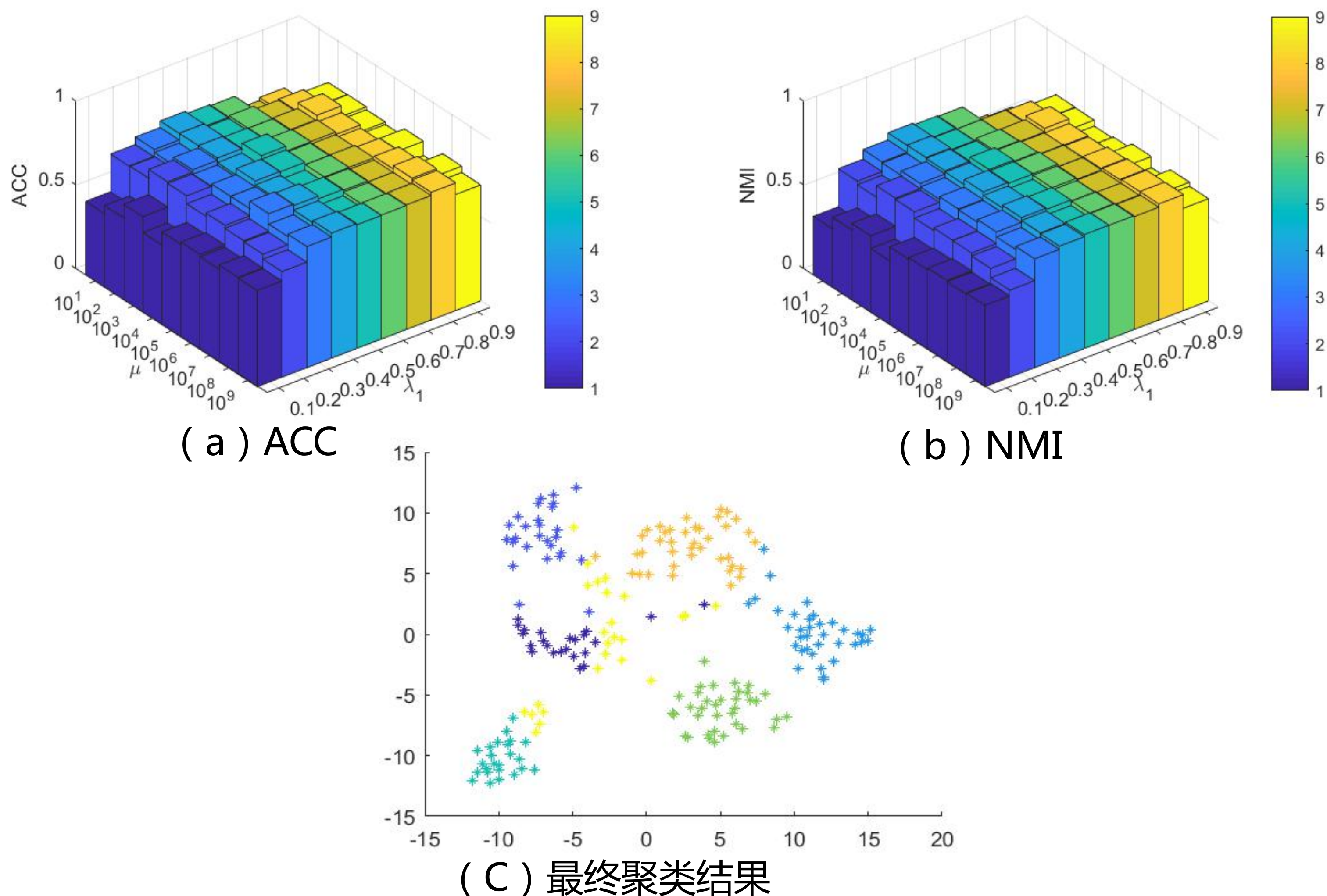


图2 MSRC_v1数据集在 λ_1 和 u 不同情况下的表现和聚类结果

参数分析

MSRC_v1数据集在和值不同情况下的表现和聚类结果。 λ_1 用于控制低秩约束， u 用于控制稀疏约束。最终聚类结果如图2(c)所示。

表1 实验结果

Method	ACC				
	BBCSport	uci-digit	3-sources	MSRC_v1	Outdoor-Scene
SC	0.425(0.012)	0.700(0.036)	0.478(0.040)	0.680(0.068)	0.180(0.001)
MCGC	0.404 (0.000)	0.676(0.000)	0.325(0.000)	0.742(0.000)	0.154(0.000)
MVGL	0.409 (0.000)	0.772(0.074)	0.307(0.000)	0.747(0.000)	0.350(0.000)
RMSC	0.566(0.000)	0.257(0.000)	0.394(0.005)	0.744(0.002)	0.240(0.000)
WMSC	0.611(0.002)	0.841(0.000)	0.424(0.005)	0.762(0.003)	0.630(0.010)
SWLRSC	0.356(0.000)	0.850(0.000)	0.704(0.000)	0.605(0.000)	0.399(0.000)
DMVSC	0.835(0.596)	0.859(0.055)	0.635(0.075)	0.776 (0.045)	0.657(0.000)

实验结果

为了避免算法随机初始化引起的误差，本次实验在数据集上均运行十次，采用“平均值(标准差)”的格式进行展示。结果如表1可见：DMVSC在数据集ACC指标上取得最优结果。相比单视角聚类算法SC，DMVSC构建的多视图结构可以更充分利用多视角信息。相比以WMSC为代表的多视角聚类算法，DMVSC利用双层结构将低秩表示和稀疏约束结合在一起，可以更全面的描述数据本身。