实验报告

一、实验内容

测试 sklearn 中以下聚类算法在 tweets 数据集上的聚类效果:

1、kmeans 算法

2、AP 近邻传播聚类算法

3、MeanShift 算法

4、spectral_clustering 算法

5、agglomerative_clustering 算法

6、DBSCAN 算法

7、GaussianMixture 算法

二、实验步骤

1、对 tweets 数据集进行预处理：分词及形成字典

2、使用 sklearn 中各种聚类算法函数处理数据

三、实验结果

四、总结

1、聚类准确率不高，需要继续调整参数以及优化程序来提高分类性能。

2、不同的算法对数据集的聚类效果不同，从 tweets 数据集上的聚类效果看：agglomerative_clusteri 算法、

GaussianMixture 算法效果较好，DBSCAN 算法较差