#### [1.Scent classification by K nearest neighbors using ion-mobility spectrometry measurements](https://www-sciencedirect-com-443.webvpn.las.ac.cn/science/article/pii/S0957417418305566)

• K最近临近算法将气味及其化学成分分类。

• 仅使用离子迁移[谱](https://www-sciencedirect-com-443.webvpn.las.ac.cn/topics/engineering/spectrometry)测量[法对](https://www-sciencedirect-com-443.webvpn.las.ac.cn/topics/engineering/spectrometry)气味/化学物质进行分类。

• 使用k维树搜索的分类大约快8倍。 （降低运算成本和算法复杂度）

•通过[主成分分析](https://www-sciencedirect-com-443.webvpn.las.ac.cn/topics/engineering/principal-component-analysis)，忽略了71–86％的特征进行分类。（① 筛除无关的特征 ② 降低维数）

* 算法原理

Xus为14维的IMS样本数据，Xus=[x1(us)…x14(us)]

Xi为训练集中的N个IMS样本，Xi=[xi,1,…xi,14]，i=1,…N

选择与样本Xus最邻近的K个同类样本作为其标签

* K的选择
  + 要在较大和较小中折中。通常，*K*为3、5或7
  + 不使用固定的K，如：[Wang等。（2006）](https://www-sciencedirect-com-443.webvpn.las.ac.cn/science/article/pii/S0957417418305566#bib0036)提出了基于统计置信度的置信最近邻规则；[程等。（2014）](https://www-sciencedirect-com-443.webvpn.las.ac.cn/science/article/pii/S0957417418305566#bib0005)一个*K*基于稀疏学习的神经网络算法；张等。（2018）提出了一种建立[决策树](https://www-sciencedirect-com-443.webvpn.las.ac.cn/topics/computer-science/decision-trees)的方法

作者使用固定的K值 数据平滑归一化

==平滑：==

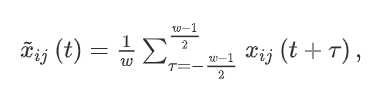


image-20200310175620593

​ 其中*x ij*（*t*）([i=1,2…N, j=1,2…14])为IMS测量值，*w*是滑动MA的窗口长度

==归一化:==

image-20200310180809741

image-20200310180809741

为归一化值，为平滑后的测量值，为均值，为标准差

* k-d树

IMS数据是14维的，为了增加精确度需要大的N，计算复杂度高必须降低*K* NN 的计算复杂度。可以使用三种不同的技术：①计算局部距离，②预先构造和编辑训练样本

③*ķ*维树（*ķ* -d树又名多维二叉搜索树）

* 新节点加入现有k-d树时，无需重新训练整个树
* 主要缺点是*k* -d树可能会错过真正的最近邻居，因为*k* -d树搜索是一种近似方法
* 对于较大的*N，*此方法通常效果很好

​ **特别适合低维，实值数据（例如IMS数据）**

* PCA降维
* ==具体步骤==

​ ① 使用PCA对离线训练集进行变换

​ 得到14通道的经验均值image-20200310212108035

​ 和包含主成分系数的14×14矩阵**C**

​ ② 对一个未知新标准化IMS采样样本进行14维的PCA变换

image-20200310213220389

image-20200310213220389

​ 其中为归一化值，为PCA变换后值。

​ ==无需对训练集进行重新训练即可添加新的训练样本==

​ ③ 对PCA-变换数据进行分类（使用KNN）

​ 计算新样本与第个经过PCA转换的训练样本**y** *i*之间的欧几里得距离。

​ 