



每个uid出现的次数总和

每个uid对应的多个prop的

种类数(频度)

简单统计

简单求和

平均值

数值最小

出现频度最高的prop

标准差

变化量的标准差

中位数

数值最大

在所有prop中进行[分类器]处理后的输出标签

频度排序取排名最高的prop值

gate二分类:有效无效

无监督学习自聚类后聚类标签

t-SNE

PCA

AP?

...

相关性

发散思维

两个prop的相关性: 构建两个prop的二维对然后聚类看聚类效果?

皮尔逊相关系数 Pearson 横向两个连续性随机变量间的相关系数

斯皮尔曼相关系数 spearman 用于衡量分类定序变量间的相关程度

肯德尔相关系数 kendall 是一种秩相关系数, 不过它所计算的对象是分类变量

已有资料