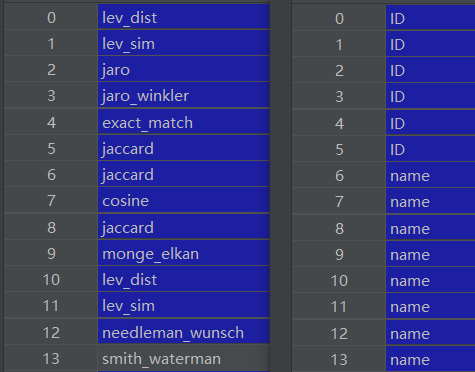
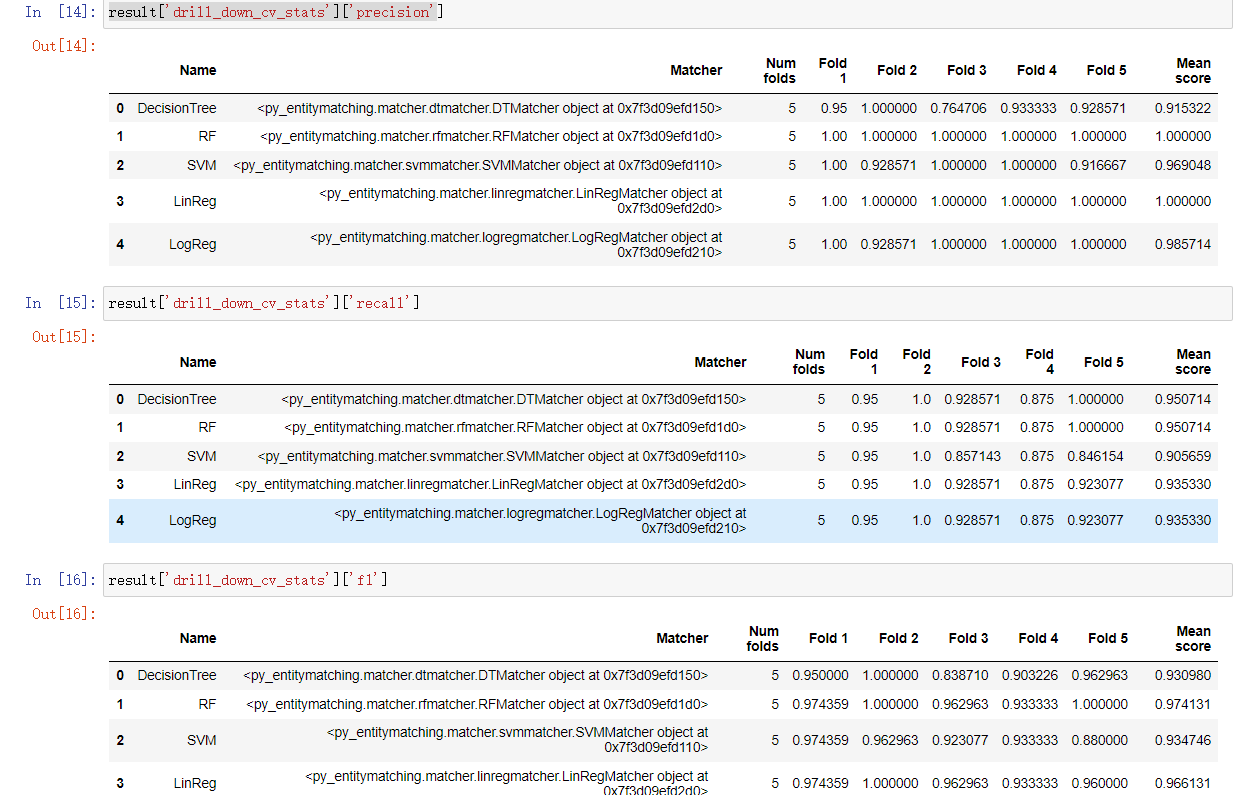
实验采用数据：

原有特征中采用了10中相似度函数，如下：

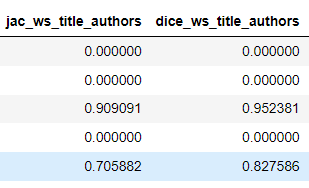


作为补充，将余下的七种相似度函数加入到特征构建中。其中，affine和汉明距离不能用于真值存在多值的情况。

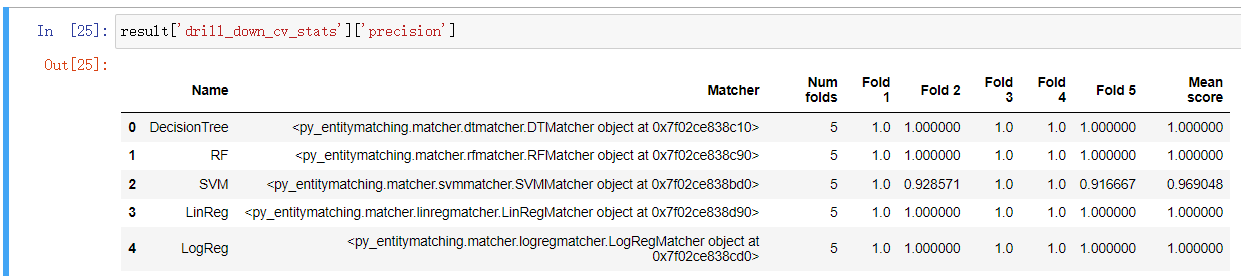
1. 不加入新特征时，现有特征输入模型的效果为：

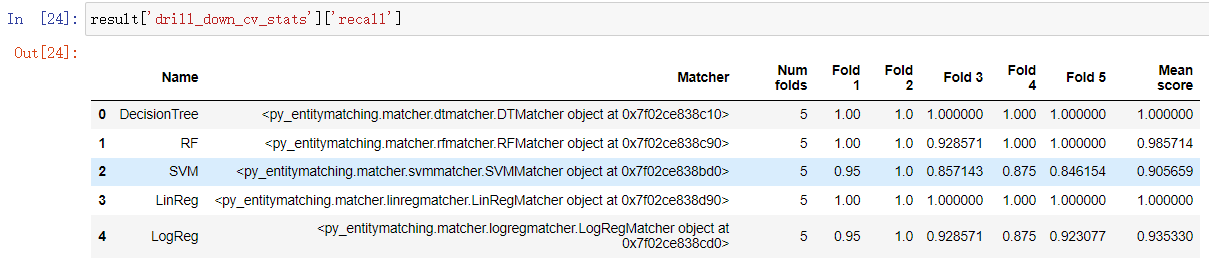


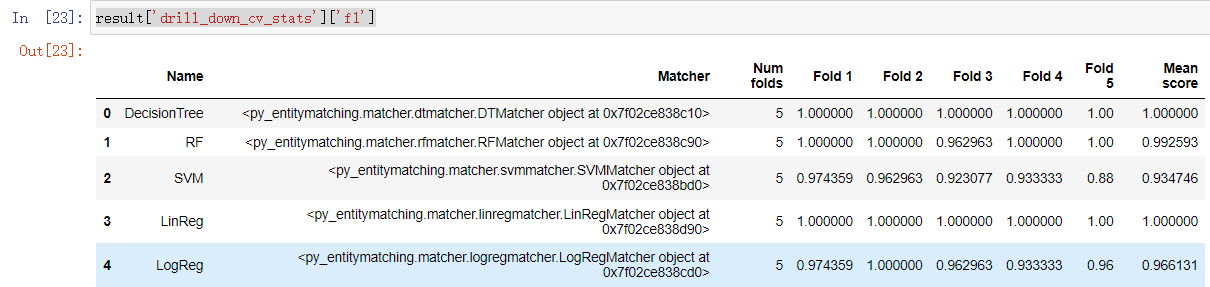
（2）首先加入dice，得到相似度数值作为特征值：



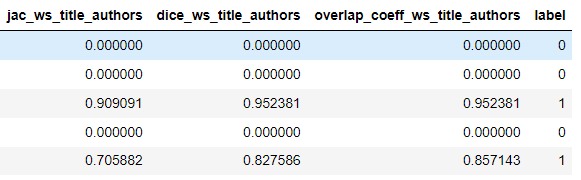
五种match模型的效果：

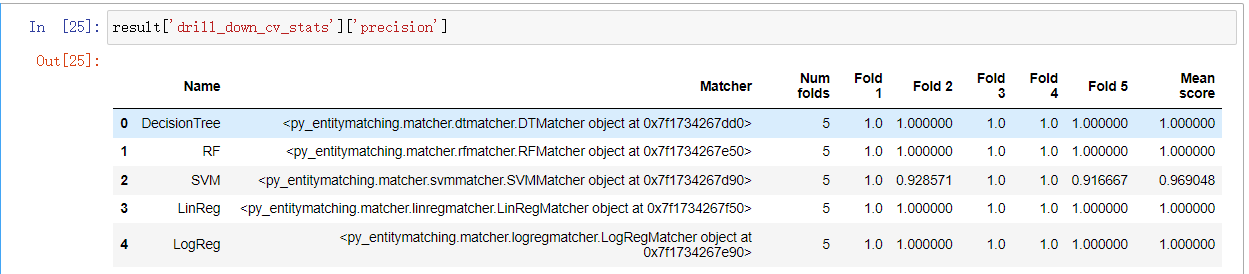


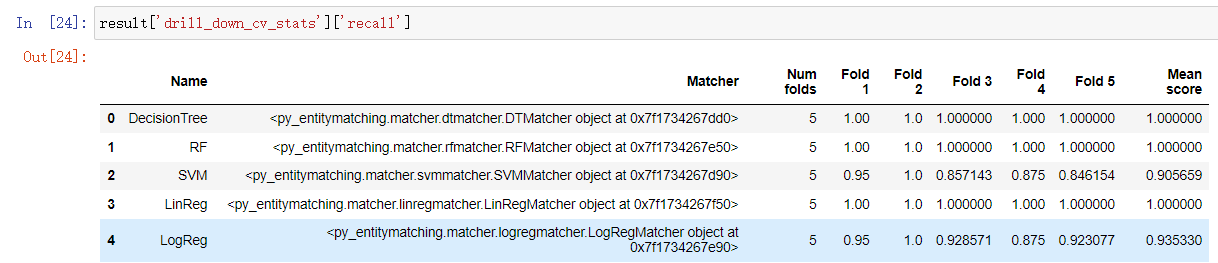


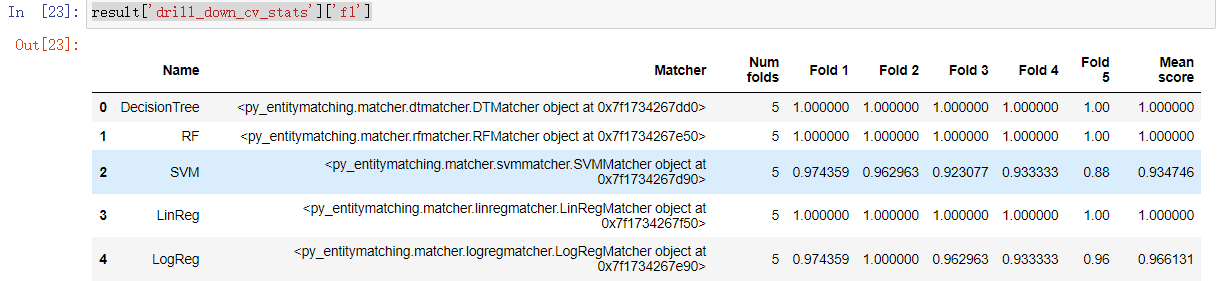


（2）再加入overlap\_coeff：

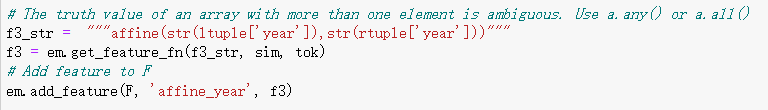








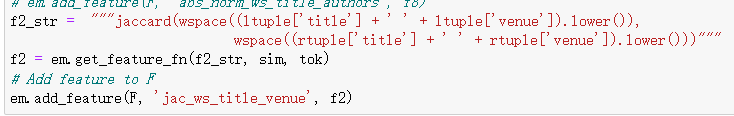
（3）加入affine函数和对rel\_diff（相关性差异）year属性计算相似度函数：



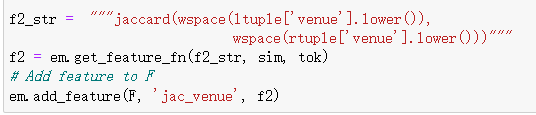
变化差异较小

（4）再次应用jaccard对venue属性（发布的会议名称）比较：

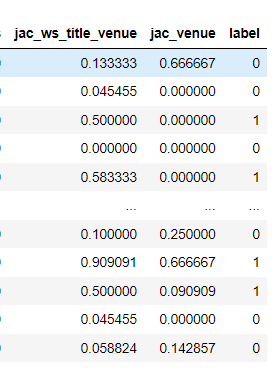
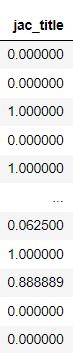
Title属性和venue联合使用



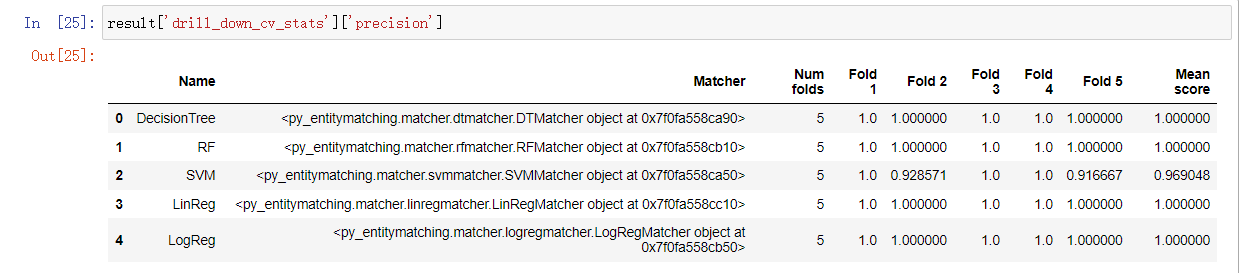
Venue单独使用：

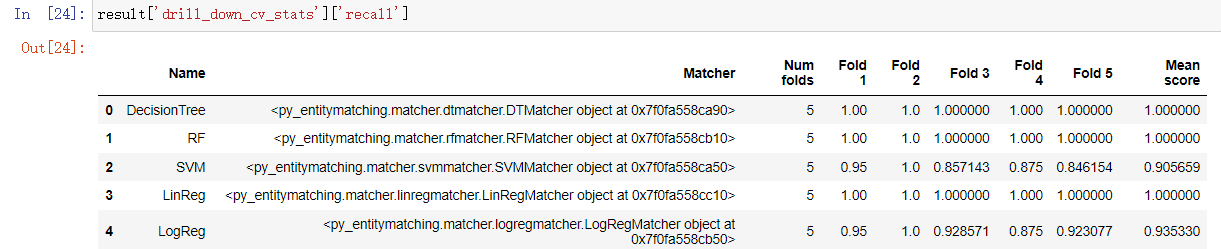


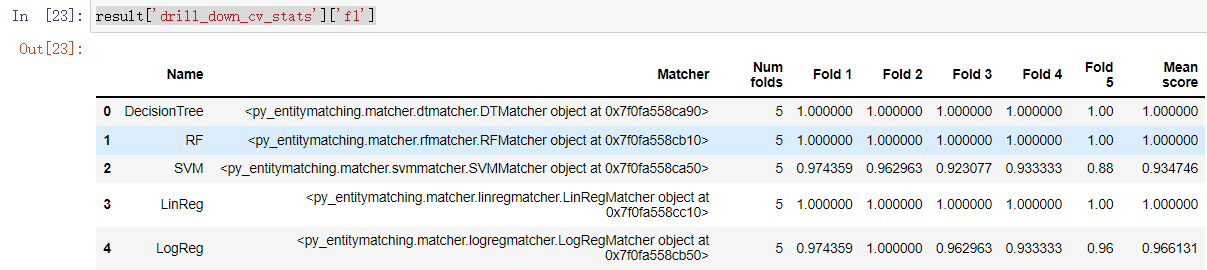
相似度：

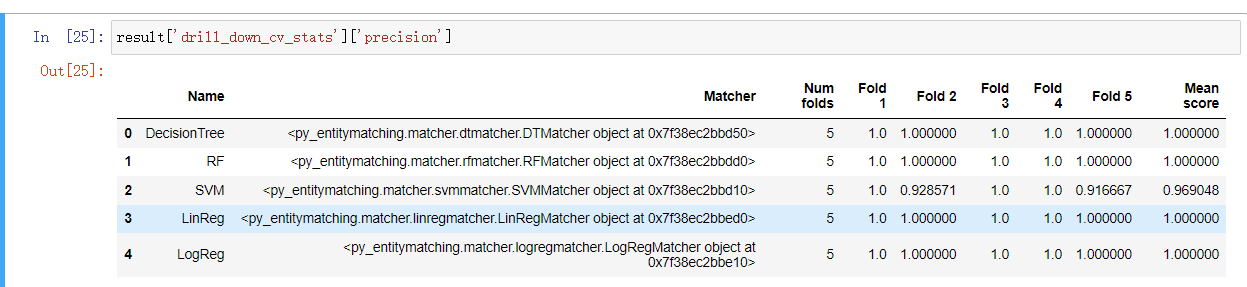
单独使用venue属性的效果：

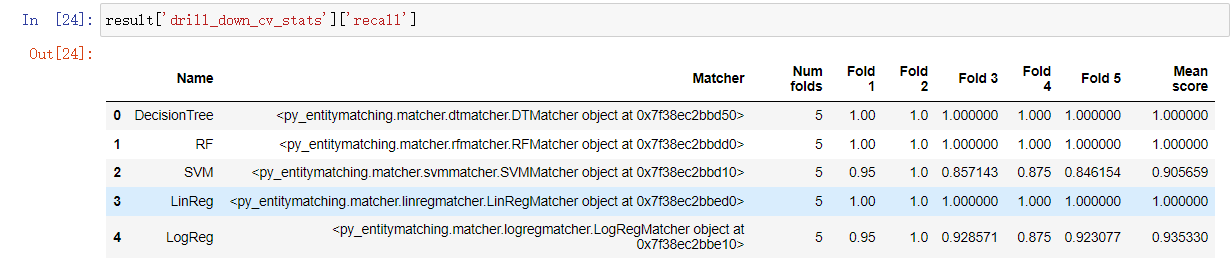


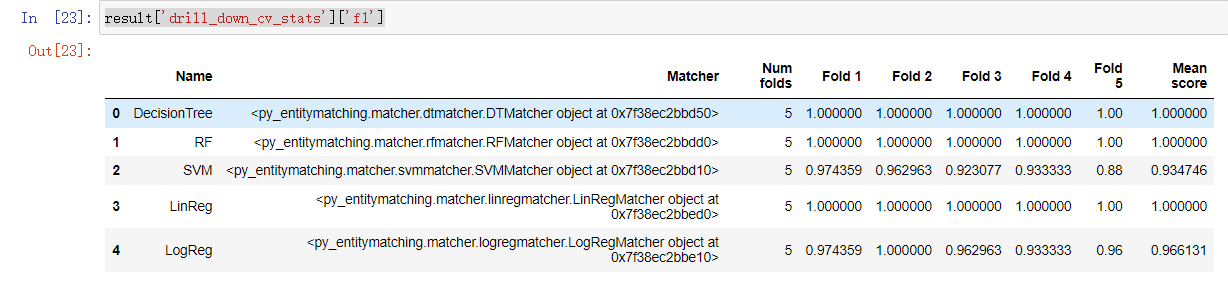




单独使用title效果：







联合使用的效果：

