**Multimodal, multitemporal, and multisource global data fusion for local climate zones classification based on ensemble learning**

2017 IEEE GRSS Data Fusion Contest参赛小组第一名小组WXYZ作品

**内容：**提出一种多模型方法对多种数据（Landsat 8数据和OSM数据）进行融合，并实现LCZ分类，最终总体精度为74.94%。

**采用数据介绍：**仅使用Landsat 8数据和OSM数据。由于Sentinel 2数据1）无长波红外波段2）每个城市只有一个时相的数据3）在某些城市有云层覆盖所以未使用Sentinel 2数据。对于Landsat 8数据，未使用赛方所提供数据，而是从Amazon S3下载的数据，分辨率包含15m、30m、100m分别对应全色波段、可见光与近红外波段、长波红外波段，所有的11个波段的数据都纳入使用。OSM数据包括“buildings”，“land-use”，“water”图层，空间分辨率为5m。

**流程：**

**1** 预处理：对影像进行辐射校正与大气校正，对全色与长波红外波段影像数据进行了归一化，使用双三性内插方法将Landsat 8所有波段数据分辨率上采样至10m，这样就可以方便地在100m空间分辨率上进行特征提取。为了减少计算量，OSM数据则降采样至10m。

**2** 特征提取：Landsat 8影像每个波段10x10（这里相当于10x10个像素合并成了100m空间分辨率的1个像素）个像素的平均值与标准差（共22个）；对Landsat 8影像的三种指数（NDVI、NDWI、BSI）进行10x10取平均值与标准差合并（共6个）；对OSM数据三个图层进行10x10取平均值处理（3个）；对NDVI与OSM“building”数据进行形态学处理得到新的特征。

**3** 分类：两种决策树集成方法——典型相关森林(Canonical Correlation Forests, CCFs)和旋转森林 (Rotation Forests, RoF)

**4** 分类后处理：对分类结果使用中值滤波进行处理，以减少椒盐噪声。最终结果通过对15个训练样本进行多数投票获得。

**结果：**总体精度74.94%，分类结果中精度较高的类别多是训练样本数量多的类别，说明通过加大训练样本数量的方式可以有效提高分类器的泛化能力。为了解决某些误分问题，可以引入数字表面模型作为新的数据源。