**Classification of Local Climate Zones Based on Multiple Earth Observation Data**

**内容：**本文采用一种基于像素的方法，介绍了Local Climate Zones (LCZ)。此外，还评估了不同的数据集与不同的分类器在一种通用框架下的性能。数据集包括：多时相热红外遥感影像，归一化数字表面模型 (NDSM)；分类器包括：支持向量机 (SVM)，神经网络，随机森林；特征选择方法：最小冗余度、最大关联法。其中利用神经网络方法与随机森林方法分别获得了97.4%和95.3的总体精度。

由于每个独特的城市类型都由各种表面材料组成，因此数据需要转换到更低的细节层次（比如从单个房屋到建筑区）。

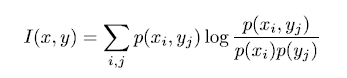
Since each specific urban class comprises various surface materials, the classified data needs to be transferred to a lower level of detail (e.g., from single houses to building blocks).

本文所关注问题：1、LCZ能否从多源地球观测数据中自动分类得到？2、分类方案能否适应欧洲的城市形态。

针对欧洲的城市形态，对LCZ分类系统进行了调整。

**特征选择：**

[互信息](https://zh.wikipedia.org/wiki/%E4%BA%92%E4%BF%A1%E6%81%AF) (Mutual Information, MI) 的概念。



**分类器：**

SVM：通过非线性映射将数据转换为一个高维空间，然后用一个最优超平面将两类划分开。对于多分类问题，则逐步进行两两分类。本文采用的是默认的线性核。

神经网络分类器：一种图，具有不同的连接层，前后层由多个节点联系（特征作为输入节点，类别作为输出节点，中间是隐含层）。每个节点都有一个加权的非线性激活函数（本文采用sigmoid函数），将输入层经过转换后传递给输出层。每个节点的权值就决定了整个网络的输出，网络采用反向传播算法进行训练。隐含层的数量默认设置为类别数量与特征数量的平均值。

The NN is a graph representing nodes (neurons) in different connected layers (with the features as input nodes, the classes as output nodes, and hidden layers in between). Each node has a weighted non-linear (here sigmoid) activation function, transferring the inputs to the output. The weights of the single nodes determine the output of the entire network and were trained by a backpropagation algorithm. The number of hidden layers was set to the mean of the class number and feature number by default. High dimensional networks construct non-linear decision boundaries of increasing complexity.

随机森林：是多个树型分类器的集成，由多个树进行投票，得到票数最多的类。

