**Week 18**

对健壮的无监督的谱特征选择的局部和全局的结构保留

**摘要**：文章提出了一种新的无监督的谱特征选择方法，来保留特征和样本的局部和全局结构。具体来说，对每一个特征，用其它特征来表示，从而保留特征的局部结构；对权值矩阵的低秩约束来保留样本和特征的全局结构；还提出了学习图矩阵来衡量样本间的相似度，从而保留样本的局部结构。

**介绍**：

之前的USFS（无监督谱特征选择）的一个共同特点是对原始数据的图矩阵的构造。首先，原始数据中有噪声和冗余，导致构建出来的图矩阵是低质量的，会降低特征选择方法的有效性。其次，在一些USFS方法中，只保留了样本的局部和全局结构中的一个。本文提出了保留两者，因为由于它们对对方信息的补偿，这两种几何结构已经被证明可以提升USFS方法的表现。第三，现有的USFS方法没有考虑特征间的相关性。最后，图矩阵的学习和特征选择是在两个分离的步骤中进行的，容易导致局部最优解。

本文贡献：

1. 利用特征水平的表达属性以及对权值矩阵的低秩约束，在特征选择中考虑了特征的局部关联和全局关联；
2. 使用了低秩约束来保留样本的全局结构，学习图矩阵来保留样本的局部结构。
3. 在本文提出的方法中，特征间的关联和样本间的关联都可以在本真低维子空间中被识别。
4. 本文在本真低维子空间中同时迭代地进行了图矩阵的构造和特征选择。

**相关工作**：

USFS方法属于嵌入式，包含两部分：图矩阵的学习来进行子空间学习，和一个产生稀疏的正则项来进行特征选择。USFS分为三类：顺序式，共同式，迭代共同式。

顺序式方法首先进行子空间学习来获得数据的图表达，然后通过产生稀疏的正则项来在图表达和原始数据之间进行一个稀疏的特征选择。

共同式在一个框架里同时进行子空间学习和稀疏特征选择。

迭代共同式认为选择到的特征高度依赖于学习到的图矩阵，因此迭代地更新图矩阵和选择到的特征，直到算法收敛。

**方法**：

3.2局部特征关联

首先对每一个特征，用全体特征表示：

 （1）

通过将对每个特征的预测看成一个任务，以及用范数来约束任务间的稀疏性，（1）式可转换成矩阵形式，因此有如下最小二乘目标函数：

 （2）

可导致行稀疏，也就是说Z的某些行为零元素，来去除X中对应的特征。

1. （2）说明每个特征由其他特征的一个子集的线性组合所表示，对应的权值向量是Z的第i列。很明显，中的值越大，对应的特征在特征的表达中就更重要。

3.3全局特征关联和全局结构保留

腐坏的数据有可能大幅增加秩，因此引入低秩约束。也就是，其中，方程(2)变为：

 （3）

几何上来说，A将X转换到新的空间，也就是通过考虑d个特征间的关联来进行子空间学习。特殊情况下，（13）式进一步证明了这样的子空间学习实际上是通过考虑全局特征的关联来实施了LDA，从而保留样本的全局结构。

3.4局部结构保留

直觉上，给定特征矩阵X和它的权值矩阵W，有如下目标函数：

 （4）

其中是高斯核函数。

方程（4）在学习W之前，从原始高维数据中学习了一个固定的图矩阵S。也就是说，图矩阵不依赖于低维子空间学习。这样的话，如果原始数据被噪声和冗余所破坏，学习到的图矩阵就是不正确的。而且方程（4）需要调整两个参数，很耗时间。特殊的，S的质量对参数很敏感。这启发我们从干净的数据中学习图矩阵，以及减少参数数目。然而图矩阵和低维子空间都不能提前获知，因此我们同时学习他们，并迭代地优化，来达到各自的最优解。于是我们可以从样本的分布来学习图矩阵，而不是利用高斯核函数，有以下目标函数：

 （5）



其中用来避免平凡解。

由于只用k个近邻来表示每一个特征，我们的方法避免了异常值的影响。

目标函数：

将（5）中的权值矩阵W看成（3）中的低秩权值矩阵AB，得到目标函数：

 （6）