# 特征（图像）分类

## 视觉特征

图像的边缘、轮廓、点、线、面、纹理等物理意义明显，易于提取的特征。

## 统计特征

灰度直方图特征、距特征（均值，方差，峰值，熵）。

## 变换系数特征

各种数学变换，如傅里叶变换，小波变换等。

## 代数特征

反映图像矩阵的某种代数属性（如特征值，特征向量等）。

# 中层特征

语义相似的patch往往在低层视觉特征上并不相似。

中层特征是对低层特征的一种聚集和整合，是对低层特征的一种概括和抽象，是一种特征的变换方式。其本质是对低层特征进行统计分布分析建立起其与语义之间的联系。应用图像场景的整体统计信息，将量化后的图像低层特征视为视觉单词，通过视觉单词分布来表达图像场景（空间结构）内容。

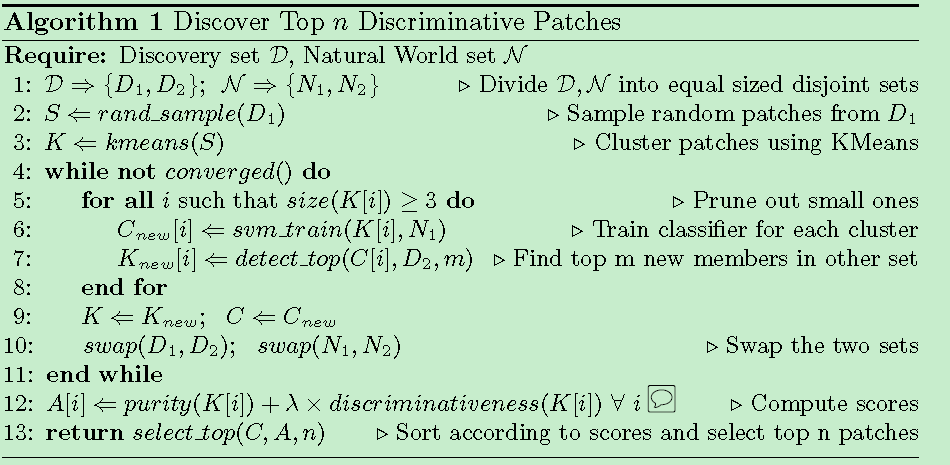
视觉单词、视觉词典、视觉词包（直方图分布）、潜在语义

词包优化

1. BOVW：Bag Of Visual Words

实现步骤：

1. 大数据聚类，找到适当的聚类中心点----Vocabulary。
2. 训练数据像聚类中心映射，得到每一个训练数据在该聚类中心空间的一个低维表示。
3. 得到每一个训练数据的低维表示后，选择适当的分类器训练。
4. 对新来的样本先映射到聚类中心空间，然后利用得到的分类器进行预测。
5. Sketch-Token:A Learned Mid-Level Representation For Contour And Object Detection
6. 通过监督分类的方法，人工勾勒图像场景结构。基于结构图像，考虑不同尺度，不同方向，自相关因素，选取特征向量并进行聚类（K-Means），得到场景结构视觉单词集合。
7. 利用随机森林的方法进行像素属于哪一类的sketch-token特征概率估计。
8. 利用临近像素特征概率求和的方法估计中心像素点属于contour的概率。
9. Usupervised Discovery of Mid-Level Discriminative Patches
10. Right primitive:patch
11. Primitive representation: bottom-up(strategy)、raw pixels(transformed-HOG 、SIFT 、DPM…)、encoding(cluster、SVM、CNN)
12. Meaning patch：representation、 discrimination
13. How to get a patch(Representation): K-means
14. Purify the patches(Discrimination): SVM classifier
15. Procedure:



# 特征选择

特征选择和特征提取都是从特征集合中选择子特征集合或者提取转换后（新的特征空间）的特征。

## 特征选择方法

1. Filter
2. Wrapped
3. Embeded

Filter方法选择特征集合中的子集作为最终的代表特征集合，其选择完全依赖于特征值集合，与具体的分类器训练过程无关。

Wrapped方法与Filter方法的最大不同点在于，该方法的特征选择过程和分类器训练过程相辅相成，该种方法的精确度更高，算法复杂度较之filter方法更高。且训练过程依赖于训练集和测试集，所以要防止过拟合问题的发生。

Embeded方法类似于Wrapped方法，不同点在于，该方法并不需要测试集合，只在训练集合上进行操作。

## 特征选择主要解决的问题

1. 特征分类准确性的评价标准
2. 特征子集搜索策略

主要的评价标准有：F-score,G-score,D-score, DFS。

其中F-score所计算的是二类特征之间的特征代表性评价指标，具体公式是类间方差之和除以类内方差之和。

G-score是F-score的多类推广。G-score存在的缺陷是其是在原始特征值下计算评价指数，未能统一特征之间的尺度和量纲。

D-score在G-score的基础上提出了离散系数，样本标准差与均值的比值，以此来统一特征之间的量纲和尺度。（scale也是一种方法，只是这个计算复杂度低一些）。

DFS方法则在D-score的基础上，考虑了特征之间的相关性（在实际操作时，是否应该先计算特征值的相关性系数，对于相关性系数高的特征进行主特征选取）。

子区搜索策略大致分为四类：

1. 顺序前向选择（Sequential Forward Selection，SFS）
2. 顺序后向剔除（Sequential Backward Selection, SBS）
3. 顺序前向浮动（Sequential Forward Floating Selection, SFFS）
4. 和顺序后向浮动策略（Sequential Backward Floating Selection, SBFS）

其中

1）初始化特征子集为空集，然后逐个搜索将评价值最大的特征加入该集合。

2）与1）相反，初始化子集集合为整个特征集合，然后搜索最小指标的特征，逐个剔除。

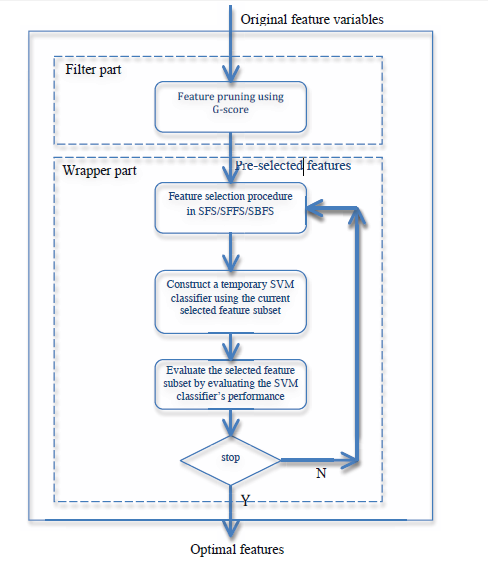
以上方法思路简单，易于操作，但是特征间存在“子集嵌套”的问题。

3）顺序前向浮动搜索方法，解决了子集嵌套的问题。与1）不同在于，在加入新的最高指标时，需要判断该指标的加入是否对准则函数（分类性能）有提高。

4）与3）类似，方向相反。

## 基于SVM的特征选择方法

1. 基于线性SVM核函数的特征权值评价方法
2. 结合fileter和wrapped的特征选取方法



# 特征提取Feature Extraction

（区别于特征转换）

## PCA方法

（PLSA\LDA）SVD分解

本质上是在寻找线性相关特征向量组的最大无关组，也即是换一个基来表示原特征空间。从而去除或者降低原特征空间的相关性和冗余度。

通过对特征的协方差矩阵（对称正定矩阵一定可以相似对角化）求特征值和特征向量的方式。可以将特征值较小的特征剔除，对特征值相同的，保留一个即可。

过程：

1. 求特征集合X的协方差矩阵C
2. 求C的特征值和对应的特征向量
3. 对特征值排列，剔除较小特征值对应的特征，对应特征值和特征向量相差不大的，选择保留一个即可

思考：

可以在计算得到特征协方差矩阵C之后，进一步求取相关性系数，这样可以观察各维度特征之间的相关性。在去相关后再行作对角化，剔除较小特征值对应特征。

## FLDA方法（Fisher线性判别分析方法）

目的也是为了降低维度，进而进行模式判别。其利用向量间内积投影的思想，将权值向量W与特征值向量X之间的内积结果作为投影后的结果y。

与SVM形式相似，但是区别在于他们所要求的最佳化目标函数不同以及判别准则不同。SVM在于同时最大化分类间隔和最小化误分类距离和。而LDA方法所采用的最佳化目标函数是要最大化类间距离之和与类内距离之和的比值。判别准则上，SVM是通过对样本在决策函数上的正负表现来判断，而LDA则根据投影后的值域和训练集选择最佳阈值作为判别界限。

过程：

* + 1. 计算样本均值向量m1、m2
    2. 计算类间离散度矩阵Sb=(m1-m2)(m1-m2)T
    3. 计算样本类内离散度矩阵和总类内离散度矩阵Sw=S1+S2
    4. 最佳w值为：w\*=Sw-1（m1-m2）

思考：

既然其形式上类似与SVM，且LDA也有类似SVM的和扩展，那么基于SVM的特征选择方法是否也可利用LDA作内核，如法炮制。。。

## PCA的核扩展

在映射后的特征空间进行PCA。推导略。

## FLDA的核扩展

在映射后的特征空间进行LDA。推导略。

# 多核学习

多核学习方法包括线性核组合方法和非线性核组合方法，是在单核SVM的基础上推广开来的方法。其基本思想是在训练SVM分类器时，其决策函数中的核函数不采用简单的单一核函数，而是采用一系列核函数组合的方式。

在机器学习时，存在一种可能是不同的特征有可能适应不同的核函数（类型和参数），也有可能某几种特征适应某一种，而另外的特征适合另一类型的核函数的可能。所以在训练时，提供一系列的备选的核函数集合，并为他们赋予权重，让核函数集合一起来决定最后的特征转换和决策函数。

相当于在原SVM分类模式中间添加了一个隐含层，如下图所示：

