# 2018.08-9中大实验室学习了解--多源信息融合与多视角学习

# 1传统与现在

传统的方法（例如PCA、SVD和LDA）对于维数比较高的数据，一般通过“硬”方法，直接降维成二维的数据形式（矩阵）进行处理。近些年，出现的2DPCA，2DSVD等新方法，是通过双向的对二维数据进行处理，在处理的过程中，不会使用“硬”的降维方法而破坏数据的结构信息。近些年来，张量数据处理的数学条件已经非常成熟，在SIAM等期刊上已经有大量的张量数据分解和处理方法出现，而这也是近些年来的一个热门方向。

张量分解的方法主要有HOSVD、HOOI以及上面提到的2DPCA和2DSVD等。上面的两种方法只限于处理三维的张量数据，具有局限性；其实也可以将它们两个看作是HOOI的一种特殊情况。相关的数学的具体文献可以再网上找到。

其应用的领域非常多。现在在模式识别和图像处理等期刊中已经有很多相关的文章出现。

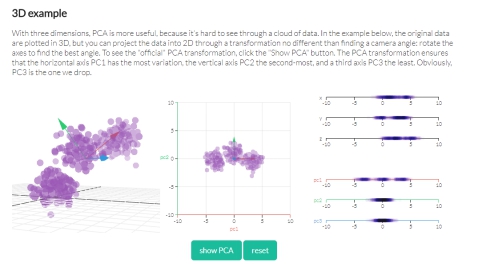
## Pca

<http://blog.csdn.net/xl890727/article/details/16898315> (matlab实现) <http://blog.codinglabs.org/articles/pca-tutorial.html> (清楚的原理解释)

### PCA算法

下面探讨一种称作主成分分析（PCA）的方法来解决部分上述问题。PCA的思想是将n维特征映射到k维上（k<n），这k维是全新的正交特征。这k维特征称为主元，是重新构造出来的k维特征，而不是简单地从n维特征中去除其余n-k维特征。

PCA算法



### 总结一下PCA的算法步骤：

设有m条n维数据。

1）将原始数据按列组成n行m列矩阵X

2）将X的每一行（代表一个属性字段）进行零均值化，即减去这一行的均值

3）求出协方差矩阵C=1/mXXT

4）求出协方差矩阵的特征值及对应的特征向量

5）将特征向量按对应特征值大小从上到下按行排列成矩阵，取前k行组成矩阵P

6）Y=PX即为降维到k维后的数据

降维问题的优化目标：将一组N维向量降为K维（K大于0，小于N），其目标是选择K个单位（模为1）正交基，使得原始数据变换到这组基上后，各字段两两间协方差为0，而字段的方差则尽可能大（在正交的约束下，取最大的K个方差）。

协方差为0，方差最大

根据上述推导，我们发现要达到优化目前，等价于将协方差矩阵对角化：即除对角线外的其它元素化为0，并且在对角线上将元素按大小从上到下排列，这样我们就达到了优化目的。

讨论：

根据上面对PCA的数学原理的解释，我们可以了解到一些PCA的能力和限制。PCA本质上是将方差最大的方向作为主要特征，并且在各个正交方向上将数据“离相关”，也就是让它们在不同正交方向上没有相关性。

因此，PCA也存在一些限制，例如它可以很好的解除线性相关，但是对于高阶相关性就没有办法了，对于存在高阶相关性的数据，可以考虑Kernel PCA，通过Kernel函数将非线性相关转为线性相关，关于这点就不展开讨论了。另外，PCA假设数据各主特征是分布在正交方向上，如果在非正交方向上存在几个方差较大的方向，PCA的效果就大打折扣了。

最后需要说明的是，PCA是一种无参数技术，也就是说面对同样的数据，如果不考虑清洗，谁来做结果都一样，没有主观参数的介入，所以PCA便于通用实现，但是本身无法个性化的优化。

## kernel PCA

代码：<http://blog.csdn.net/wsj998689aa/article/details/40398777>

而对于所研究对象主成分(或主元)是非线性的情况下，PCA就失去了有效性。在种情况下，kernel PCA就应运而生了。并且kernel PCA是目前比较流行的方法。

KPCA的算法虽然简单，但是个人认为，它的意义更在于一种思想：将数据隐式映射到高维线性可分空间，利用核函数进行处理，无需知道映射函数的具体形式。这种思想实在是太牛了，它让降维变得更有意义。为这种思想点赞！！！

## SVD

SVD和PCA呢，都实现了降维与重构，但是呢，思路不太一样，

，线性代数里也叫作奇异值分解，可以用来计算矩阵的特征值（奇异值）和特征向量，SVD大多数情况下用来提取一个矩阵的主特征值，可以用于图像的压缩和去噪

代码：1936-用100个奇异值

1. Im=imread('niu.jpg');
2. figure(1);
3. Im=rgb2gray(Im);
4. imshow(Im,[]);
5. set(gca,'position',[0 0 1 1])
6. [m n]=size(Im);
7. Im=double(Im);
8. r=rank(Im);
9. [s v d]=svd(Im);
10. Im2=s(:,:)\*v(:,1:100)\*d(:,1:100)';
11. figure(2);
12. imshow(mat2gray(Im2));
13. imwrite(mat2gray(Im2),'1.jpg')
14. set(gca,'position',[0 0 1 1])

## ICA信号处理

实例代码： <http://blog.csdn.net/yeyang911/article/details/48733963>

<http://blog.csdn.net/cai2016/article/details/52983473>

博客1 <http://blog.csdn.net/cai2016/article/details/52983473> （容易理解--做ppt）理解

<http://blog.csdn.net/neal1991/article/details/45128193>

代码：<http://blog.csdn.net/SMUEvian/article/details/70244493?locationNum=8&fps=1>

述流程图中，由独立源线性组合成的观测信号X，只需要对S进行求解即可。S=B\*X，其中B为A的逆，通过迭代对A进行逼近，当达到设置的精度时即可分离出混合矩阵的近似。盲源分离的目的是求得源信号的最佳估计。

### 非高斯分布（讨论

那么ICA在本质上是不可能实现的。假定S经过混合矩阵A后，他们的联合概率密度仍然不变化，因此我们没有办法在混合中的得到混合矩阵的信息。

### 步骤

ICA（独立成分分析）（继续学习看书签）  
高斯分布：正态分布  
ICA估计方法：(即对条件假设的描述：目标函数，依据它进行无监督学习)  
1）非高斯最大化（负熵、高阶累积量--常用四阶累积量）；  
2）互信息最小化；  
3）最大似然估计；  
4）KL散度；  
确定目标函数之后，采用一定的算法（各种自适应优化算法）寻优处理。  
应用：  
在闹磁图（MEG）中分离非自然信号、在金融数据中找到隐藏的因素、自然图像中减少噪声、人脸识别、图像分离、语音信号处理、远程通信等  
ICA处理的详细步骤如下：  
1）零均值：中心化：也就是求x均值，然后让所有x减去均值，这一步与PCA一致。  
2）白化：通过线性变换把随机变量转换成相互无关的变量  
3）ICA:

### 白化

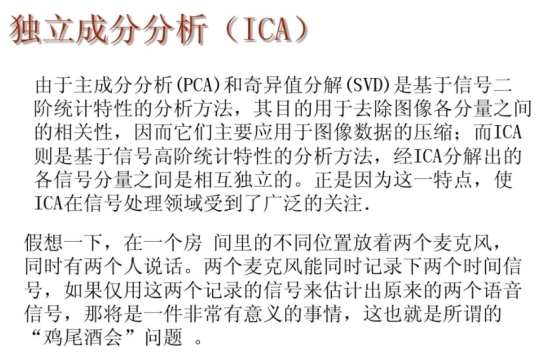
白化分为PCA白化、ZCA白化

<http://blog.csdn.net/haoji007/article/details/52790189>

<http://ufldl.stanford.edu/wiki/index.php/%E7%99%BD%E5%8C%96>

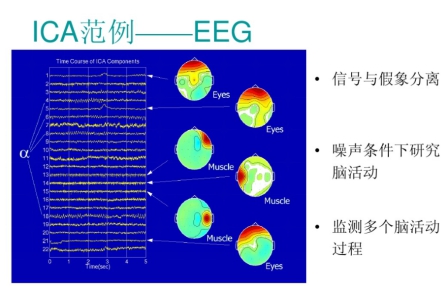
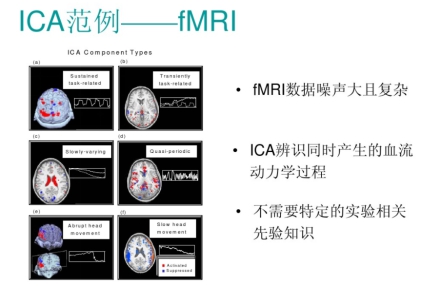
pca预处理过程称为**白化**（一些文献中也叫**sphering**）。举例来说，假设训练数据是图像，由于图像中相邻像素之间具有很强的相关性，所以用于训练时输入是冗余的。白化的目的就是降低输入的冗余性；更正式的说，我们希望通过白化过程使得学习算法的输入具有如下性质：(i)特征之间相关性较低；(ii)所有特征具有相同的方差

**理解**：ZCA 白化是一种数据预处理方法，它将数据从 C:\Users\sereny1\AppData\Local\Temp\ksohtml\wps5C5C.tmp.jpg 映射到 C:\Users\sereny1\AppData\Local\Temp\ksohtml\wps5C5D.tmp.jpg 。 事实证明这也是一种生物眼睛(视网膜)处理图像的粗糙模型。具体而言，当你的眼睛感知图像时，由于一幅图像中相邻的部分在亮度上十分相关，大多数临近的“像素”在眼中被感知为相近的值。因此，如果人眼需要分别传输每个像素值（通过视觉神经）到大脑中，会非常不划算。取而代之的是，视网膜进行一个与ZCA中相似的去相关操作 (这是由视网膜上的ON-型和OFF-型光感受器细胞将光信号转变为神经信号完成的)。由此得到对输入图像的更低冗余的表示，并将它传输到大脑。



### 应用领域（相关论文）





机器学习系列博客整理

<http://blog.csdn.net/u013802188/article/details/40903471>

# pca和ICA的区别（讨论）

PCA attempts to find uncorrelated sources, where as ICA attempts to find independent sources.

简单讲：PCA是一个降维的过程，ICA则是帮助你从多个维度分离有用数据的过程。

附上之前看过的两个链接很不错：

PCA: [Principal Component Analysis explained visually](https://link.zhihu.com/?target=http%3A//setosa.io/ev/principal-component-analysis/)

ICA: [What is the difference between PCA and ICA?](https://link.zhihu.com/?target=https%3A//www.quora.com/What-is-the-difference-between-PCA-and-ICA)

2

在PCA中，您想要找到的基础是最能说明数据变异性的基础。 PCA基础的第一个向量是最好地解释数据的可变性（主方向）的第二个向量，第二个向量是第二个最佳解释，并且必须与第一个向量正交。

在ICA中，您想要找到的基础是每个向量是数据的独立组件，您可以将数据视为混合信号，然后ICA基础将具有每个独立信号的向量。

（Both PCA and ICA try to find a set of vectors, a basis, for the data. So you can write any point (vector) in your data as a linear combination of the basis.

In PCA the basis you want to find is the one that best explains the variability of your data. The first vector of the PCA basis is the one that best explains the variability of your data (the principal direction) the second vector is the 2nd best explanation and must be orthogonal to the first one, etc.

In ICA the basis you want to find is the one in which each vector is an independent component of your data, you can think of your data as a mix of signals and then the ICA basis will have a vector for each independent signal.

）

3

当您想要找***到数据的降级表示***时，PCA有帮助

当您想要将**数据的表示形式视为独立的子元素**时，ICA有帮助。

以外行人士称PCA有助于压缩数据，ICA有助于分离数据。

注意：**PCA和SVD是相同的，**通常更好的是使用中心数据矩阵的SVD，因为SVD算法比PCA更快和数字更稳定。

注2：在某些情况下，**NMF（非负矩阵分解）可以作为ICA工作。** 在NMF中，您希望找到的基础是帮助您将数据重建为基础向量的正相加。 这意味着基础将包含代表原始数据部分的向量，如果您的数据包含图像，则NMF基础包含可帮助您重建数据集中任何图像的图像部分。

# 信息融合

**信息融合**(information fusion）起初被称为数据融合

### 参考博客

<http://blog.csdn.net/congzhao27/article/details/11105787>

### 主要方法

利用多个传感器所获取的关于对象和环境全面、完整信息，主要体现在**融合算法**上。因此，多传感器系统的核心问题是选择合适的融合[算法](http://baike.baidu.com/view/7420.htm" \t "_blank)。对于多传感器系统来说，信息具有多样性和复杂性，因此，对信息融合方法的基本要求是具有[鲁棒性](http://baike.baidu.com/view/45520.htm" \t "_blank)和并行处理能力。此外，还有方法的运算速度和精度；与前续预处理系统和后续信息识别系统的接口性能；与不同技术和方法的协调能力；对信息样本的要求等。一般情况下，基于[非线性](http://baike.baidu.com/view/392135.htm" \t "_blank)的[数学方法](http://baike.baidu.com/view/96884.htm" \t "_blank)，如果它具有容错性、自适应性、联想记忆和并行处理能力，则都可以用来作为融合方法。多传感器数据融合虽然未形成完整的理论体系和有效的融合算法，但在不少应用领域根据各自的具体应用背景，已经提出了许多成熟并且有效的融合方法。

主要算法

### 主要算法

多传感器数据融合的常用方法基本上可概括为随机和[人工智能](http://baike.baidu.com/view/2949.htm" \t "_blank)两大类,随机类算法有加权平均法、[卡尔曼滤波](http://baike.baidu.com/view/302394.htm" \t "_blank)法、多[贝叶斯估计](http://baike.baidu.com/view/6960491.htm" \t "_blank)法、[证据推理](http://baike.baidu.com/view/4494735.htm" \t "_blank)、产生式规则等；而人工智能类则有[模糊逻辑](http://baike.baidu.com/view/338152.htm" \t "_blank)理论、神经网络、粗集理论、[专家系统](http://baike.baidu.com/view/10875.htm" \t "_blank)等。可以预见，神经网络和人工智能等新概念、新技术在多传感器数据融合中将起到越来越重要的作用。

### 应用

信息融合技术目前在军事、模式识别、遥感图像、医学图像处理等领域，具有广泛的理论研究意义。

**信息融合又分为像素级融合、特征级融合、决策级融合**。

本文主要概述了特征级融合技术的**基本概念、几种主要方法以及融合质量的评估方法**。概述

**信息融合**可以描述为：综合多源的信息，得到高品质的有用的信息。各种单一的传感器往往不能从场景中提取足够的信息，以至于很难甚至无法独立获得对一副场景的全面描述。需要**多传感器的同时获取目标数据进行融合分析**，才可有效的进行分类识别决策。多传感器信息融合又分为**像素级融合、特征级融合、决策级融合**三个层次。

## 特征级融合

是从各个传感器获取的信息中提取目标特征，然后通过融合算法获取融合目标特征量，进行目标分类识别。特征级融合目标识别，既能保持足够数量的目标有效信息，去除冗余信息，又提高了目标识别的精确性。

本文以特征级融合研究作为**重点**。近几年特征级融合方法开始受到人们的重视，并且取得了许多成果。

现在特征级融合**目标**就是要通过某种算法将n个传感器获取的所有特征融合为高品质的有用的信息，然后将得到的融合特征作为进一步决策的依据。

### 特征级融合方法：大致分为五大类

概率论统计方法、逻辑推理方法、神经网络方法、基于特征抽取的融合方法和基于搜索的融合方法

#### ****概率统计方法****

**2.1加权平均法**

权重可看成不同传感器准确性的度量。该方法简单直观，但是必须事先对各个传感器进行详细的分析，获取它的权重。并且在不同特征维度上每个传感器的准确性都不一样，所以**权重的获取**成为主要难点。

**2.2贝叶斯估计法**

Bayes方法在用于多传感器信息融合时，是将多传感器提供的各种不确定信息表示为概率，并利用概率论中Bayes条件概率公式对他们进行处理的一种方法

假设有n个传感器，分别对感兴趣区域进行特征提取后，对于某一维度的特征值表示为x1,x2…xn; 假设融合后可能的特征值为M，由x1,x2…xn得到M的概率为p（M|x1，x2）,依据贝叶斯准则有：。

在先验概率已知的情况下，贝叶斯准则是最佳的融合准则，可给出精确融合结果。但是在实际应用中，各个传感器很难获得所需的先验概率，这一点大大限制贝叶斯准则的应用。

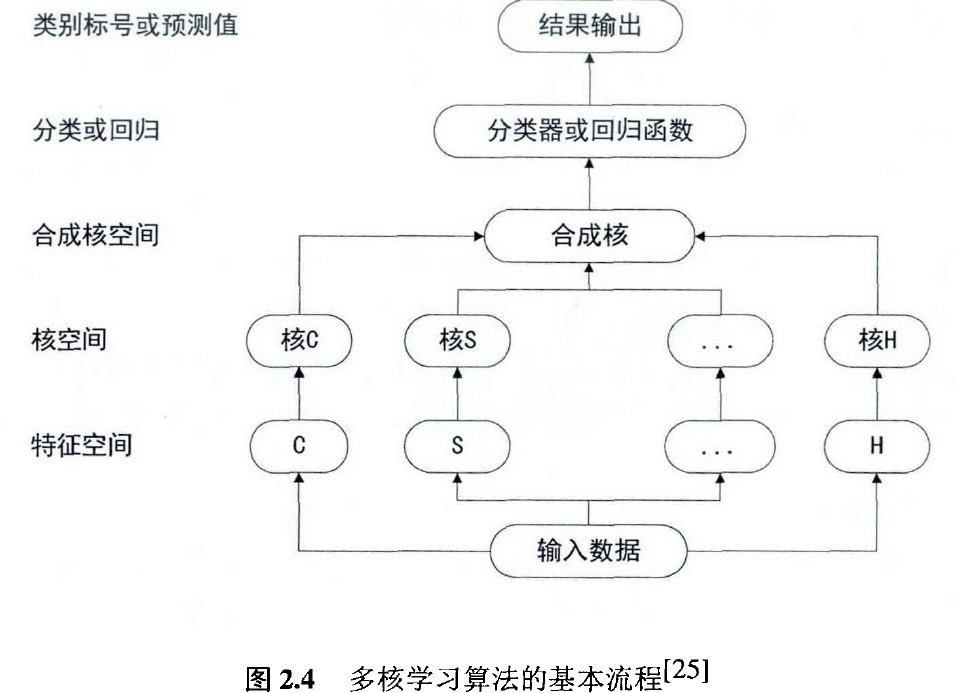
# 多视角学习9.9

## 多核学习的研究

多核模型是一类灵活性更强的基于核的学习模型, 在多核框架下, 样本在特征空间中的表示问题转化成为基本核与权系数的选择问题。

**研究人员大都研究**（1）核函数权系数的选择问题。如非平稳的多核学习方法, 局部多核学习方法,非稀疏多核学习方法等（2）多核学习理论。如早期的基于Boosting的多核组合模型学习方法, 基于半定规划(Semide¯nite programming, SDP) 的多核学习方法, 基于二次约束型二次规划(Quadratically constrained quadratic program, QCQP)的学习方法, 基于半无限线性规划(Semi-in¯nitelinear program, SILP)的学习方法, 基于超核(Hyperkernels)的学习方法, 以及近来出现的简单多核学习(Simple MKL)方法和基于分组Lasso思想的多核学习方法。

**多核学习的算法流程**



**多核学习缺点**：

尽管多核学习在解决一些异构数据集问题上表现出了非常优秀的性能，但不得不说效率是多核学习发展的最大瓶颈。首先，空间方面，多核学习算法由于需要计算各个核矩阵对应的核组合系数，需要多个核矩阵共同参加运算。也就是说，多个核矩阵需要同时存储在内存中，如果样本的个数过多，那么核矩阵的维数也会非常大，如果核的个数也很多，这无疑会占用很大的内存空间。其次，时间方面，传统的求解核组合参数的方法即是转化为SDP优化问题求解，而求解SDP问题需要使用内点法，非常耗费时间，尽管后续的一些改进算法能在耗费的时间上有所减少，但依然不能有效的降低时间复杂度。高耗的时间和空间复杂度是导致多核学习算法不能广泛应用的一个重要原因

## 多核学习在图像分类中的应用

<http://baidutech.blog.51cto.com/4114344/742984/>

在刚用SVM做分类的时候对一个现象一直比较困惑，看到大家**将各种不同类型特征，拼接在一起，组成庞大的高维特征向量，送给SVM，**得到想要的分类准确率，一直不明白这些特征中，到底是哪些特征在起作用，哪些特征组合在一起才是最佳效果，也不明白为啥这些特征就能够直接拼在一起，是否有更好的拼接方式？后来了解到**核函数以及多核学习的一些思想**，临时抱佛脚看了点，对上面的疑问也能够作一定解释，正好拿来和大家一起探讨探讨，也望大家多多指点。本文探讨的问题所列举的实例主要是围绕项目中的**图像分类展开，涉及SVM在分类问题中的特征融合问题。**扩展开来对其他类型分类问题，理论上也适用。

**特征融合：**主要用来描述各种不同的特征融合方式，常见的方式有前期融合，就是前面所描述的将各个特征拼接在一起，后期融合本文后面会提到

**核函数：**SVM遇到线性不可分问题时，可以通过核函数将向量映射到高维空间，在高维空间线性可分

**多核学习：**在利用SVM进行训练时，会涉及核函数的选择问题，譬如线性核，rbf核等等，多核即为融合几种不同的核来训练。

#### （1）不好的方法

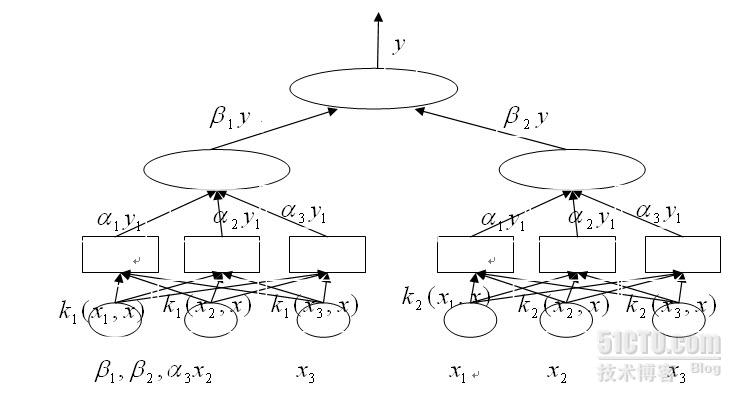
1. 你所提取的所有特征，全部串在一起，一定合适么？如果我想知道哪些特征组合在一起效果很好，该怎么办？  
2. 用svm进行学习时，不同的特征最适合的核函数可能不一样，那我全部特征向量串在一起，我该如何选择核函数呢？  
3. 参数的选取。不同的特征即使使用相同的核，可能最适合的参数也不一样，那么如何解决呢？  
4. 全部特征都计算，计算时间的花销也是挺大的

对于刚才的问题，如果用前期融合，可能是用下面方式来解决：  
1. 根据经验，觉得在样本中可能表现不错的特征加进来，至于组合么，全部串在一起，或者选几个靠谱的串一起，慢慢试验，慢慢调，看哪些特征有改进就融合在一起  
2. 也是**根据经验，选取普遍表现不错的RBF核**，总之结果应该不会差  
3. 交叉验证是用来干嘛的？验证调优参数呗，全部特征融合在一起，再来调，**尽管验证时间长，不要紧，反正模型是离线训练的**，多调会也没关系。

#### （2）好的方法

多核学习(MKL)可能是个不错的选择，该方法属于**后期融合**的一种，通过对不同的特征采取不同的核，**对不同的参数组成多个核，然后训练每个核的权重，选出最佳核函数组合**来进行分类。

Svm的分类函数形似上是类似于一个神经网络，输出由中间若干节点的线性组合构成，而多核学习的分类函数则类似于一个比svm更高一级的神经网络，其输出即为中间一层核函数的输出的线性组合。其示意图如下：

[](http://img1.51cto.com/attachment/201112/203229112.jpg)

在上图中，左图为普通SVM示例，而全图则为MKL示例。其中

#### 准备工作：

1. 人工标注地图类别正负样本，本次标注正样本176张，负样本296张  
2. 提取正负训练样本图片的A~G各个特征  
3. 归一化特征  
4. 为每个特征配置对应的核函数，以及参数

#### 7个人经验与总结：

1多核学习在解释性上比传统svm要强。多核学习可以明显的看到各个子核中**哪些核在起作用**，哪些核在一起合作效果比较好。  
2. 关于参数优化。曾经做过实验，关于同一特征选用同一核，但是不同参数，组合成多个核，也可以提升分类准确率。  
3. 多核学习相比前期特征融合在性能上会有3%~5%左右的提升。  
4. 通过特征选择，可以节约特征计算时间。

工具：

Shogun工具盒：<http://www.shogun-toolbox.org/>，其中关于该工具的下载，安装，使用实例都有详细说明。该工具除了提供多核学习接口之外，几乎包含所有机器学习的工具，而且有多种语言源码，非常方便使用。

# 多视角实验室论文

## Flexible and Robust Multi-Network Clustering

### 内容

#### 摘要：允许跨不同网络进行多个底层群集结构的框架

propose a flexible and robust framework that allows multiple underlying clustering structures across different networks.Our method models the domain similarity as a network,which can be utilized to regularize the clustering structures in different networks. We refer to such a data model as a network of networks (NoN). We develop NoNClus, a novel method based on non-negative matrix factorization (NMF),to cluster an NoN. We provide rigorous theoretical analysis of NoNClus in terms of its correctness, convergence and complexity. Extensive experimental results on synthetic and real-life datasets show the effectiveness of our method

提出了一个灵活可靠的框架，允许跨不同网络的多个基础聚类结构。我们的方法将域相似性建模为网络，可以用于规范不**同网络中的聚类结构。** 我们将这样的数据模型称为网络网络（NoN）。 我们开发了NoNClus，**一种基于非负矩阵分解（NMF）**的新方法，聚集一个NoN。 我们在其正确性，收敛性和复杂性方面对NoNClus提供严格的理论分析。 综合和现实的数据集的广泛的实验结果显示了我们的方法的有效性

#### 关键词：网络聚类，图聚类

**主要：a robust and flexible multi-network clustering method**

#### 内容：

我们调查新型网络设置NoN中的新的聚类问题，其中多个底层聚类结构可以在域特定网络之间共存。 它概括了现有的**多视图和多域聚类方法的单一聚类结构假设**，并且在许多新兴的现实生活问题中具有更广泛的适用性。

•我们开发了一个新型的两阶段聚类框架，NoNClus，可以从主要网络的指导下同时聚类域特定网络，工作。 NoNClus允许跨不同大小的域特定网络进行部分映射，这比多视图设置更为通用和逼真。 我们也对NoNClus的正确性，收敛性和复杂性进行严格的理论分析。

现有的多网络聚类方法主要用于多视图网络[37,20,19]。在多视图聚类中，视图可以是网络[37,20,19]或数据特征矩阵[1，23，35]。集群聚类[30,12,13]与多视图聚类相关，其中通过在不同视图上应用相同的聚类算法或应用不同的聚类算法获得共聚群集

#### 过程实战？

有效性评估

模拟研究

20新闻组数据集

绩效评估

组织特异性基因共表达网络的案例研究

#### 4主要的网络聚类

#### 1非负矩阵分解（NMF）

#### 2谱聚类

#### 3运用到实际中，数学太空

4正确性分析,收敛分析,复杂性分析

## Graph clustering

#### 图形聚类的定义和方法，

在图中找到“相关”顶点集合。 我们回顾了图中什么是集**群和集群质量测量**的许多定义。 然后，我们提供用于为输入图的**整个顶点集生成聚类的全局算法**，之后我们讨论通过本地计算识别特定种子顶点的聚类的任务。 给出了关于图聚类算法应用领域的一些想法。 我们还讨论了评估集群和基准集群算法的问题。

图形是由一组顶点（也称为节点）和一组边缘组成的结构，这些边是顶点对之间的连接。**图形聚类**是将图的顶点分组成簇的任务，考虑到图形的边缘结构，使得每个聚类之间应该存在许多边缘，并且在集群之间相对较少。在将给定输入图的顶点分组为聚类的意义上的图形聚类，这是本次调查的主题，不应该与基于结构相似性的图集集聚混淆

### 熟悉术语

### 学习：计算复杂度，近似算法，图论和马尔可夫链。

## 孙仕亮- PAC-Bayes Analysis of Multi-view Learning

本文提出了八个PAC-Bayes边界来分析多视图分类器的泛化性能。这些边界采用数据依赖的高斯先验，强调具有高视图协议的分类器。前两个界限的先前的中心是原点，而第三和第四边界的先前的中心由数据相关向量给出。获得这些边界的一个重要技术是两个导出的对数行列式不等式，其差异在于是否涉及数据的维数。第五和第六边界的中心是在训练集的单独子集上计算的。最后两个界限使用未标记的数据来表示视图协议，因此适用于半监督多视图学习。我们评估所有呈现的多视图PAC-Bayes边界的基准数据，并将其与以前的单视图PAC-Bayes边界进行比较。讨论了多视图边界的有用性和性能。

## Multi-view Learning Overview: Recent Progress and New Challenges

### 多视角学习

是机器学习的一个新兴方向，它考虑**学习多个视角**来提高泛化性能。多视图学习也被称为数据融合或来自多个特征集的数据集成。自2013年初进行多视图机器学习的最后一次调查以来，多视角学习近年来取得了长足的进步和发展，面临新的挑战。本概述首先回顾理解基础，了解多视图学习的属性和行为。

然后，多视角学习方法以三个类别进行描述，以提供整洁的分类和组织。对于每个类别，提出了代表性算法和新提出的算法。

本次调查的主要特点是在早期方法的一致性的基础上，综合介绍了多视角学习方法的最新发展

例如，对于图像和视频，颜色信息和纹理信息是两种不同的特征，可以被视为双视图数据。在网页分类中，通常有两个视图描述给定的网页：网页本身的文本内容和锚文本链接到此网页的任何网页

2013年初，多视角学习取得长足进展。无论从多方面利用数据信息还是从机器学习分支应用的角度出发，新提出的多视点学习方法在一定程度上都具有优势。这些多视角学习方法也可能激发方法学研究和实际应用。因此，有必要介绍多视角学习的最新发展，分析其特点以及有希望的应用。与以前的综述相比，本文的内容和结构是全新的。首先，在早期方法的一致性的基础上，综合介绍多视角学习方法的最新发展。此外，为了显示多视图学习方法的清晰结构，通过从较高级别的新类别的分类来总结多视图学习方法。

2为了了解多视图学习方法的特点和性能，先后提出了一些基于PAC-Bayes理论和Rademacher复杂度理论的泛化误差分析。 这里我们介绍两种最近提出的泛化误差分析，PAC-Bayes边界和Rademacher复杂度的泛化误差界限

最后，我们将讨论几个最新的开放性问题和挑战，为未来的研究提供有希望的场地。

具体来说，在本文中，**多视角学习方法分为三大主要类别**

#### 3. Multi-view Learning Method

**co-training style algorithms, co-regularization style algorithms and marginconsistency style algorithms.**

类别：共同训练风格算法，协整规则风格算法和边际一致性风格算法。

1)Co-training style algorithms are enlightened byco-training[2]. **Co-training** is one of the earliest methods for multi-view learning for which learners are trained alternately on two distinct views with confident labels for the unlabeled data. For example, co-EM [3], co-testing [4], and robust co-training [5] belong to this co-training style algorithm.

2) For **co-regularization** style algorithms, the disagreement between the discriminant or regression functions of two views is regarded as a regularization term in the objective function. Sparse multi-view SVMs [6], multi-view TSVMs [7],multi-view Laplacian SVMs [8] and multi-view Laplacian TSVMs [9] are representative algorithms.

3) Besides the two conventional style algorithms**, margin-consistency style algorithms ar**e recently proposed to make use of the latent consistency of classification results from multiple views [10, 11, 12, 13]. They are realized under the framework of maximize entropy discrimination (MED) [14]. Different from the co-regularization style algorithms which make restrictions on the discriminant or regression functions from mul-

多视图传输学习、多视图聚类、多视点判别分析，多视点半监督学习和多任务多视图学习。

#### 4. Multi-view Datasets and Software Packages

为了提供多视角学习研究的实验支持，我们

描述一些广泛使用的多视图数据集和代表性的软件包。

手写数字数据集1手写数字数据集。二进制图像数字化的每个数字（类）共有200个示例（共2,000个示例），在该数据集中以六个特征集（视图）表示。

广告数据集2广告数据[10]包括3,279个例子，包括459个广告图像（正面例子）和2,820个非广告图像（负面例子）。第一个视图描述图像本身（图像的URL中的单词，替代文本和标题），而另一个视图包含所有其他功能（包含图像和图像指向的页面的URL中的单词）。

WebKB数据集3 WebKB数据[10]包括从四所大学的计算机科学系网站收集的1,051个双视图网页：康奈尔大学，华盛顿大学，威斯康星大学和德克萨斯大学。有230个课程页面和821个非课程页面。

1数据可在http://archive.ics.uci.edu/ml/datasets/Multiple+Features获得。

2数据可在https://archive.ics.uci.edu/ml/datasets/Internet+Advertisements获得。

3数据可在http://www.cs.cmu.edu/afs/cs/project/theo-11/www/wwkb/获得。

4 Data are available at https://datahub.io/dataset/multipie.

5 Data are available at http://mmlab.ie.cuhk.edu.hk/archive/cufsf/.

6 Data are available at http://www.cbsr.ia.ac.cn/english/HFB%20Databases.asp.

7 Data are available at https://archive.ics.uci.edu/ml/datasets/Corel+Image+Features.

8 Both available at http://www.cst.ecnu.edu.cn/ slsun/software/MUDAcode.rar.

9 Available at <http://www.davidroihardoon.com/code.html>.

具体介绍看论文

#### 开放性问题

1大规模多视图学习

2深度学习和多视角

多视图深度学习  
  
  
对于多视图学习，通过并入多视图学习算法和深度学习方法，还有一些提高性能的潜力。迄今为止，多视图深度表示学习有两个主要策略[130]。首先，Ngiam等人[131]提出通过从被认为是分割自动编码器的测试时间可用的视图重建两个视图来提取共享表示。其次，Andrew等[132]提出了CCA的DNN扩展称为深CCA。对于实际应用，Zhu et al。 [133]提出了一种多视角感知器，它是学习面部身份和视图表示的深层模型。苏等人[134]提出了一种新颖的CNN架构，其将来自3D形状的多个视图的信息组合成单个且紧凑的形状描述符。 Elhoseiny等[135]通过在CNN中采用视图不变表示，实现了多视图数据的联合对象分类和姿态估计。 Elkahky等人[136]提出了一个总体推荐框架，它使用深度学习来将丰富的用户特征与项目特征相匹配。他们还展示了如何扩展这个框架，以组合来自不同领域的数据，以进一步提高推荐质量。虽然这些方法已经在多视角学习框架中实现了深入的学习，但在方法和应用方面仍然有很多开发多视图深度学习的空间。例如，多视图深度高斯处理是一种有趣且具有挑战性的模型。

5.3。模型设计超过两个视图

基于两个视图提出了许多多视图学习算法。其实，

多视图数据在实际应用中非常普遍。用于处理多个视图的一些现有方法是双视图方法的变体。它们通过目标函数中的加法运算来组合所有的成对相关性。然而，该策略的主要缺点是只探索了特征对之间的相关信息，忽略了高阶统计。除了以上简单的组合双视图学习算法的方法之外，还有一些新的

提出处理多视察案件的策略。其中，张量产品是用于学习多视图数据的有效技术。提出的TCCA [85]是使用张量产品到CCA的有效实例。如何开发更加丰富的多视角学习算法与张量产品是值得研究的问题。此外，考虑到多视图学习方法的多样性，在特定设置下为两个以上视图设计算法将是有趣的。

5.4。多视图学习与不完整的意见

现有的多视图学习算法在不同应用中已经表现出有希望的性能。这些算法通常在全视图假设下工作，其中需要观察所有视图的数据。在一些实际应用中，这种全视图设置可能会被违反。例如，由于传感器故障或人为错误，某些某些视图的数据可能会丢失。因此，我们只能访问视图不完整的多视图数据，这对多视图学习带来困难。在视图不完整的情况下，如何良好地执行多视角倾斜算法，或提出具有处理不完全视图的能力的新的多视图学习算法是一个有趣的研究方向。通过驾驶视图不完整的实际应用，开发了一些关于不完整视图学习的工作。大多数情况下，它被设计为处理特定任务，如多视图聚类[137,138]。不完整视图的多视图聚类的主要思想是使用空间变换重建丢失视图中的数据，然后执行完整的多视图学习方法。由于贝叶斯方法可以通

5.5。基于高斯过程的多视图主动学习

。由于高斯过程主动学习[141,142]已经被提出并被实验证明是有效的，多视图高斯过程可以是一个优雅的贝叶斯学习方法，值得研究基于高斯过程的多视图主动学习算法

5 .6. 贝叶斯框架下的多视图顺序模型