# 浏览式阅读

## 1 自己的总结、评价以及应用

Multi-view clustering via CCA：

Here, we consider constructing such projections using multiple views of the data, via Canonical Correlation Analysis (CCA).

（传统的数据聚类问题通过将高维数据映射到低维数据空中，然后对数据进行聚类操作。在multi-view数据聚类问题中，我们使用CCA对multi-view数据进行聚类操作）

我们的算法优点：

我们表明，算法成功所需的分离条件比以前的结果弱得多。换言之。我们的算法实现聚类的条件要比以前的算法的要求简单很多。

## 2 文章的主要问题（abstract、疑问句中）

## 3 结论（abstract以及conclusion中）

## 4 思路脉络（小标题中的关键句）

1. **Introduction**

（下面的有点看不太懂）

一个术语：a mixture model

有一个基本问题basic problem：

We are given independent samples from a mixture of k distributions，下面是我们的基本任务：

1. 推导出相关属性infer properties of the underlying mixture model
2. 对某一个样本判断它所属的分类 classify a random sample according to which distribution in the mixture it was generated from

针对这个问题，最近的工作焦点在“没有sepration”或者sepration尽可能少的前提下对数据进行分类。（聚类分离的条件很重很重要，换言之，就是用哪一个指标作为分类的依据，需要多少指标才合适？）

上面的工作自然促成了一个假设natural multi-view assumption：

That the views are (conditionally) uncorrelated, conditioned on which mixture component generated the views.（这个假设是接下来的研究前提）

基于上面的假设，我们提出了一个基于CCA的子空间学习方法subspace learning method：

提出这样一个方法的原因：we are able to (approximately) find the low-dimensional subspace spanned by the means of the component distributions.当不同视角的数据映射到子空间时，数据分布的方式得到了很好的分离。在子空间中，属于同一分布的数据点之间的距离就会很近，比原空间中分布距离要近得多。（这就是将数据映射到一个子空间的目的所在。）

最后，实验结果也表明我们的方法要比PCA主成分分析的方法的表现效果好很多。

### Related Work

我们的工作借鉴了两个算法：

### **This Work**

我们的输入数据源：

对于数据源我们有两个条件：

1. we require that conditioned on the source, the two views are uncorrelated
2. 第二个要求简单而言就是： there is sufficient correlation（有足够的相关性） between the views

后边巴拉巴拉说了一大堆，就是说基于CCA的多视图聚类要比PCA的各种好。。。　。。。

1. The Setting

基本看不懂，公式概念太多了。。。　。。。

1. The Clustering Algorithm

首先介绍一个引理lemma，然后基于这个引理介绍我们的算法

Algorithm 1.

①将S随机分成两个字迹A和B

②Let Σb12(A) (Σb12(B) resp.) denote the empirical covariance matrix between views 1 and 2； Compute the top k − 1 left singular vectors of Σb12(A) (Σb12(B) resp.), and project the samples in B (A resp.) on the subspace spanned by these vectors.

③Apply single linkage clustering [DE04] (for mixtures of log-concave distributions), or the algorithm in Section 3.5 of [AK05] (for mixtures of Gaussians) on the projected examples in View 1.

下面介绍了两个定理 theorem ：

①Theorem 1 (Gaussians)

②Theorem 2 (Log-concave Distributions)

1. Analyzing Our Algorithm

一大堆公式、证明与推理，看不懂。。。　。。。

1. Experiments

介绍了在两种数据集上的实验表现，主要比较对象是PCA

* 1. Audio-visual speaker clustering

5.2. Clustering Wikipedia articles

## 5 难理解点