

指导教师： 杨涛

提交时间： 2015.3.29

The task of
Digital Image Processing

数字图像处理

School of Computer Science

No: 1

姓名： 张玉杉

学号： 2012302420

班号： 10011202

人的动作在背景中无监督光谱的双分配聚类

西蒙·琼斯，凌少

电子电气工程系

设菲尔德大学，设菲尔德，S1 3JD，英国

邮箱: simon.m.jones@sheffield.ac.uk

ling.shao@sheffield.ac.uk

摘要

一个最近的研究趋势表明背景信息怎么和动作联系起来，如场景或对象，可以增强人体动作识别系统的准确性。然而，使用背景来提高无监督人的行动聚类是前所未有的，不能使用现有的聚类算法实现。为了解决这个问题，我们介绍一种新的，通用的算法双重任务 k -均值 (DAKM)，这是唯一能够进行合作产生的聚类任务的同时，利用相关信息提高聚类。此外，我们为更好地表现现实数据描述一个 DAKM 频谱扩展 (SDAKM)。大量的实验（合成数据和三个真人动作数据集包括现场背景）表明 DAKM/SDAKM 比国家做先进的聚类算法更好（考虑到动作和场景之间的关系）。

1.介绍

最近许多在计算机视觉领域的研究致力于从不同的来源表示和识别人的行为，如 YouTube 视频和好莱坞的 LMS。在这些现实的视频，人的行为尤其应当考虑相当大的背景，动作执行的地方或对象。这方面的信息可以被集成到一个动作识别系统来帮助消除相似的类之间的歧义，从而提高分类结果，如 Marszałek et al 【7】。如果一个动作的背景是一个篮球场，这告诉我们这个动作更可能是指打篮球图 1 说明了这个概念



图 1：上图：从视频各种剧照，显示出人类的行为发生在不同的场景
下图：一个动作场景的关系表的简单模型的例子。此表中的信息可以从监督学习训练数据直接推断，但在无监督学习必须与动作场景类别同时估计。

场景（有监督动作）的识别已经被探究，然儿，还没有实验研究被考虑用在无监督人的动作的场景聚类。这是一个重要的考虑，准确的无监督和半监督聚类是人类行为的许多实际工作的关键，如自动标注视频数据库和基于内容的快速视频检索。行为聚类也可协助两视频

内容的场景相似性判断，例如，可用于提高视频数据库的推荐系统。但是，这不是简单的应用现有的行动的背景下，研究行为的聚类分析。在现有的工作中，标记的训练数据通常适用于动作种类与场景之间的关系，它允许直接推断他们的场景-它是简单的构造基于这种关系的场景识别模型。为对行动和场景的聚类的目标，另一方面，没有提供参照的标签，所以行动/场景关系无法直接获知。而是要同时估计行为聚类，场景聚类，以及动作/场景关系。

在本文中，进行无监督行动与背景的聚类下，我们提出了双分配聚类思想和新的的双重任务的 Kmeans 聚类算法 (DAKM)。这两个数据集的聚类算法能根据数据集的两种观点之间关系，以提高聚类。我们首先证明合成数据 DAKM 理论的适用性，然后结合谱表示显示在几个现实的人类行为的数据集的最先进的结果 (用行动和场景的两个视图)。

这项工作的其余部分如下，我们概述了聚类和场景中的人体动作识别部分以前的作品 2。3 定义的双分配聚类，和细节的双重任务 k-均值 (DAKM) 聚类算法以及其频谱扩展。实验的综合数据和三个真人动作的数据集已经给出 4。我们的结论与我们的研究探讨在 5。

2. 以前的工作

最近的许多人的动作的识别实验已经考虑到场景的影响。Marszalek et al.[7]证明了两个分类可以一个动作，一个场景，然后结合使用，提高与一组相关联的两分类权重的分类结果。ikizlercinbis 和 sclaro[3]更进一步，结合对象，在多实例学习框架的场景和行动的信息，提高了分类性能的 YouTube 视频。Prest et al. [10]使用弱监督框架来学习人类的行为和场景中的物体之间的相互作用，特别是学习行为和对象之间的空间关系。然而，所有这些技术都依赖于实验数据，学习行为与场景之间的关系。其他的工作都集中在人的行为完全无监督聚类。Yang et al. [14]证明了一个全球性的行动描述和时间匹配算法基于局部特征的方法提供了优秀的聚类结果。Niebles et al. [9]使用 PLSA 和 LDA 技术起源于自然语言处理，来聚类行动基于中间的主题与他们有关的。Wang et al. [13]表明，谱聚类有效性，利用线性规划技术找到行动的图像对之间的距离。

近年来在通用数据聚类的发展中也应考虑。特别是，多视图聚类 [1, 4] 和 替代聚类 [2] 承担一些相似的双分配聚类。我们设想怎样将这些概念和我们自己的比较在图 2 中。

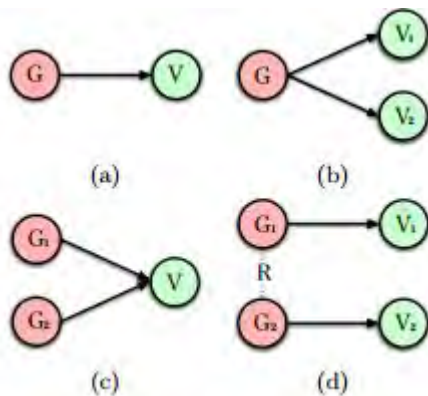


图 2：一个可视化的各种聚类方法，示出的数据集的分类方法之间的依赖关系，一个可视化的各种聚类方法，显示出的数据集的分类方法之间的依赖关系，潜在的分类方法数据集 G 和观察视图数据集 V 。(a) 普通聚类，(b) 多视点两视图聚类，(c) 两种选择的替代聚类，(d) 双分配聚类。

多视图聚类使用相同数据的多个视图，而不只是一个视图，来提高聚类性能。它假定有一个单一的，真实的数据集的聚类，视图之间的相互信息可以用来找到这个聚类。

选择聚类假定一个单一的数据集的聚类有多个有效的解决方案。然后发现这些多个聚类解决方案，基于数据集的单一视图，最大化每个聚类的优点，并且所有的聚类之间的相异性

和正交性野最大化。在我们的算法中，双分配聚类，我们假设有两有效数据集的聚类（类似替代聚类）但我们也有两个数据视图（类似于多视图聚类）。每一个有效的聚类有一个视图关联。我们估计这两种聚类之间的交互信息，可以同时提高两者的聚类结果。

3.双分配聚类

在本节中，我们定义的双分配聚类问题，描述一个可能的多目标优化方法，然后描述的双分配 k-均值算法（DAKM）作为这个优化逼近。

3.1 定义

我们定义特定的双分配聚类问题如下，我们希望聚类一组视频成离散组相类似的视频。我们假设有两个独立的视频，有效的聚类：第一个是基于视频的场景；第二聚类是根据视频的动作。我们还假设这两个视频聚类是不互相独立的如果场景是未知的，他提供了动作在这个场景发生的概率。最后，我们假设每个视频有两个视图-一个视图（来自运动功能）是由视频的动作产生的，而另一种（来自静态特征）是通过对视频场景生成。目的是产生一个动作场景聚类和一个场景聚类，估计动作和场景之间的关系，提高了求解精度。

在现实的情况下，动作和场景之间是多对多的关系。一个场景，可以与多个动作相关（例如，骑自行车和遛狗可以在公园里发生的），一个单一的动作可以与多个场景相关（例如，打篮球无论是在室内或室外都可以进行）。此外，在现实的数据集的某些（动作、场景对）是比别的更有可能，某些组合是不可能的（例如，在游泳池里打篮球）。我们想要抓住这复杂的多对多关系，不同于在多视图聚类隐式的一对一的假设（一个动作对应于一个场景）。

我们可以使用相关矩阵 $R \times R$ 的关系模型，捕捉相关信息，利用边缘联合概率分布：

$$R \equiv \frac{p(A, S)}{p(A)p(S)} \quad (1)$$

如果动作和场景的标签是已知的，的联合分布 $P(a; S)$ 可以使用相对列联表 F 近似，其中的每个条目 $F_{a;s}$ 表示包含动作和场景的视频数据集 S $P(A)$ 和 $P(S)$ 被计算为在整个数据集的动作和场景的相对边际的频率，分别为 M_A 和 M_S 的代表。 R 是这样计算的：

$$R = F \oslash (M_a \otimes M_s) \quad (2)$$

\otimes 和 \oslash 表示 Kronecker 积和 Hadamard 除法， $R_{a,s} = 1$ 表示 a 和 s 不相关。

大于一表示正相关，小于一表示负相关。

3.2 优化问题

为了进一步了解聚类问题，我们定义它为一个优化问题。首先，给定一组观测值 $(x_1; x_2; \dots; x_N)$ ，硬聚类算法优化后的基本 k-均值：

$$\arg \min_C \sum_{i=1}^n \sum_{j=1}^k C_{i,j} \|x_i - \mu_j\|^2 \quad (3)$$

在硬 k-均值聚类， C 是一个二元聚类隶属度矩阵，其中每个元素 $C_{i,j}$

表示观察 x_i 是否属于第 j 个聚类，每次观察只属于一个聚类。 μ_j 是第 j 个聚类的中心。双分配问题，目标是聚类相关的两套观测（或视图）。 $(x_1; x_2; \dots; x_N)$ 和 $(Y_1; Y_2; \dots; Y_N)$ ，把 k^x 和 k^y 分别设成 C^x 和 C^y ，相对应的 x_i 和 y_i 共发生。这是

一个未知的（但不为零）的关系。我们提出了一个修改到原来的 k-均值问题，在 C^x 和 C^y 使多目标优化问题。第一个目标函数是：

$$\arg \min_{C^x, C^y} \sum_{i=1}^n \sum_{j=1}^{k^x} \sum_{l=1}^{k^y} C_{i,j}^x \|x_i - \mu_j^x\|^2 C_{i,l}^y \|y_i - \mu_l^y\|^2 \quad (4)$$

这个目标于原来的 k-均值问题本质上是相同的，延长两套观测。它的目的是减少距离和聚类质心 X 和 Y 。这个目的需要两个距离的东西而不是全部东西，我们不需要考虑 X 和 Y 之间的任何规模的变化。第二个目标函数是：

$$\arg \min_{C^x, C^y} - \sum_{j=1}^{k^x} \sum_{l=1}^{k^y} R_{j,l} \log(R_{j,l}) \quad (5)$$

R 是通过公式 2 计算使用 C^x 和 C^y 。方程 5 大致对应 C^x 和 C^y 之间的最大交互信息。

我们计入这个目标是因为我们对 x 和 y 之间聚类的一个稀疏关系感兴趣。方程 5 是最大时， R 是一个统一的矩阵，两个聚类之间不共享信息，既然这样，我们宁愿用两个独立的聚类来提高时间效率。 $H(R)$ (X, Y 之间的联合熵) 减小， R 近似一个一一对应关系（或当 $k^x \neq k^y$ 时一个多对一关系）在 x 和 y 之间的聚类，大量的信息在这两个聚类之间共享。

如果 $H(R)$ 太低，然而， R 可能因噪音扭曲，或太稀疏，而无法准确地代表 X 和 Y 之间的关系，所以我们权衡它使用第一个目标。为了平衡这两个目标我们下面详细的正则化方法。

3.3 双分配 k-均值算法 (DAKM)

简单的期望最大化聚类最初似乎是一个直接优化目标上理想的解决方案。这是类似于用于通过替代聚类最小化交互信息的方法 [2]。这种方法是棘手的，但是，由于我们的第二目的介绍每一个 C^x 和 C^y 之间的依赖性，所以，我们设计了一个迭代更新方案优化目标方程 4 和 5。完整的方法如算法 1 所示：

算法 1：双分配 k-均值算法 (DAKM)

数据： 两组观察数据 $(x_1; x_2; \dots; x_n)$ 和 $(y_1; y_2; \dots; y_n)$, x_i 和 y_i 共发生

k^x 和 k^y 每个数据集的聚类数 λ 参数控制 R 最终稀疏

结果： 隶属度向量 C^x 和 C^y

```

begin
  (Cx, μx) ← Kmeans (x, kx)
  (Cy, μy) ← Kmeans (y, ky)
  repeat
    R ← UpdateRelationships(Cx, Cy, λ)
    (Cx, Cy) ←
      UpdateMemberships(x, y, μx, μy, R)
    μx ← UpdateCentroids(x, Cx, kx)
    μy ← UpdateCentroids(y, Cy, ky)
  until Cx and Cy don't change
Function UpdateRelationships(Cx, Cy, λ)
  R ← zeroes(kx, ky), Mx ← zeroes(kx)
  My ← zeroes(ky)
  for i ← 1 to n do
    RCix, Ciy ← RCix, Ciy +  $\frac{1}{n}$ 
    MCixx ← MCixx +  $\frac{1}{n}$ 
    MCiyy ← MCiyy +  $\frac{1}{n}$ 
  R ← R ⊙ (Mx ⊗ My) (Eqn 2)
  R ←  $\frac{\log(1+\lambda R)}{\sum_{\mathbb{R}} \log(1+\lambda R)}$  (Eqn 6)
  return R
Function UpdateMemberships(x, y, μx, μy, R)
  for i ← 1 to n do
    for j ← 1 to kx do
      for l ← 1 to ky do
        dists(j, l) ←  $\frac{\|x_i - \mu_j^x\| + \|y_i - \mu_l^y\|}{R_{j,l}}$ 
      (Cix, Ciy) ← arg minj,l dists
  return (Cx, Cy)
Function UpdateCentroids(d, C, k)
  for i ← 1 to k do
    μi ← mean({d | (d, c) ∈ {(d, C)} ∧ c = i})
  return μ

```

首先，这两个数据集的聚类成员的初始化分别使用原来的 k-均值算法（或用另一种聚类算法有更好的保证，如 k-均值++）。然后 R, C^x, C^y, μ^x 和 μ^y 更新迭代的过程中还受到原 k-均值影响。

第一次迭代是估计 R。我们根据方程式 2 计算 R，然后将它回复正常，所有的元素变成一个。

然后通过方程 6 得到 R'：

$$R'_{j,l} = \frac{\log(1 + \lambda R_{j,l})}{\sum_{j,l} \log(1 + \lambda R_{j,l})} \quad (6)$$

λ 是一个用户定义的控制 R' 和 λ ≥ 1。

第二步迭代更新的变量 C^x 和 C^y。对每一对 (x_i, y_i)，我们计算每一对聚类之间的距离，j ∈ [1..k^x], l ∈ [1..k^y]，然后除以 R'_{j,l}，然后我们找最有价值的（最小的）j 和 l 如下：

$$\arg \min_{j,l} \frac{\|x_i - \mu_j^x\| + \|y_i - \mu_l^y\|}{R'_{j,l}}$$

R' 是除数，更多的点趋向于归属频繁的聚类对，少部分点归属于频率低的聚类对。经过多次迭代，这隐含地减少第二目标有效的 R-熵如公式 5 所示。最后一步是为了更新聚类使所有的数据集独立。这是为了是他们的普通的 k-均值相同。当 C^x 和 C^y 不变时算法终止。

作为一个正则化参数——它是平衡的双目标优化问题。为了更高值的 λ ， R' 需要趋向于更高的熵，更均匀分布，和更高的重量放在距离的聚类中心最小化的地方；为了更小值的 λ ， R' 趋向于更小值的熵，或者更稀疏，和把更多的中心放在数据集的交互信息上。在实际使用中，一个较低的 λ 更有效但对噪音的抗性低，因此 λ 在噪音数据上的值要高，或者当 x 和 y 之间关系复杂（ λ 值过低），效率会降低（低于初始基本线），但是，低值 λ 会产生一个更好的聚类方案，当数据集是干净和 R 较稀疏的时候。有两个缺点 DAKM 可以在今后的工作中改进。首先，DAKM 的计算复杂度每次迭代为 $O(k^x k^y n)$ ，其中 N 是在数据集项目数。例如，DAKM 是最实用的低值 k^x 和 k^y ，

DAKM 的一个分层的扩展可以被认为是更大的聚类数的更合适的选择。第二，DAKM 只能近似于上面给出的目标函数的功能，它不可发现适合每个人的局部最优解。在一个扩展的后期工作我们希望展示一种改性 DAKM 收敛到一个单一的，正规化的目标函数。尽管如此，在 DAKM 下的实验可靠的终止了，有少于 100 次迭代的好结果和我们展示的一贯的卓越的精确度。

3.4 谱表示

该方法扩展了原来的 k-均值算法。然而，k-均值并不总是对自然的数据产生最佳聚类结果-特别是当集群对不是线性可分的。这在人类行动的分析中是明显的，Yang et al. [14] 使用谱聚类和 d Nibbles et al. [9] 使用自然语言技术 LDA 和 PLSA，而不是选择 k-均值。为了给我们的算法取得谱聚类的优势，我们采用了 DAKM 来结合双分配聚类和谱聚类。首先，让我们观察，步骤 5 的谱聚类算法 Ng et al. [8] 是普通 k-均值聚类，一次，执行算法中的步骤 1 至 4，[8] 中是数据集中的两种视图，用 DAKM 来替换第 5 步来利用两个谱表示之间的共同信息。我们称这种修改算法光谱 DAKM (SDAKM)，我们将其应用到人类行为的聚类分析 4.2。

4. 实验

在这一节中我们详细介绍几种聚类实验。比较实验结果的真实，我们使用的聚类的准确性度量提供在 [14]。如果每个聚类 c 包含数据点 $x_1; \dots; x_n$ ，每个数据点与地面真确的标签联系 $l_1; \dots; l_n$ ，聚类 c 的标签 l_c 为：

$$\arg \max_{l_c} \sum_{i=1}^n \begin{cases} 1 & \text{if } l_c = l_i \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

然后我们计算具有相同的标签分配给他们的整个数据集的聚类数据的百分比。

4.1 合成数据聚类

我们创造展示 DAKM 适用几种合成数据集。

4.1.1 合成数据的生成

为生成的人工数据集，我们首先设置 X 和 Y 总簇数，($k^x = 12, k^y = 8$)。为了获得 X 和 Y 之间的关系，我们随机产生的地面真实的关系矩阵 \mathbf{R}_{ground} ，其中 1 项表明两集群可以共同出现，而 0 项表示相反。我们保证每行的中间缓存至少有一个正的条目。在 X 和 Y 的每个簇由一个二维高斯分布表示，其中的意思是随机选择一个范围值之间的协方差矩阵，对角矩阵，和协方差的条目变化之间的范围内的值。然后，产生 10000 个样本，每个样本，然后 x 和 y 产生的两个聚类按 \mathbf{R}_{ground} 同时选择，然后两向量产生高斯聚类分布。这些合成的数据允许我们测试如何影响各种数据集的属性特别是 DAKM 的性能，重点研究的影响：1) \mathbf{R}_{ground} 中的关系数量；2) 集群分离做得如何。我们比较结果与普通 k-均值聚类，所有的合成实验，我们设置 $\lambda = 1$ 。

4.1.2 合成结果

我们首先看 x 和 y 之间不同关系的数量的影响（通过控制 \mathbf{R}_{ground} 的非零项的数目实现）。我们在普通 k-均值聚类 and 聚类性能 DAKM 之间在一个范围内的值 $|\mathbf{R}_{ground}|$ 在图 3 中的合成数据集的差分。正如预期的那样，随着 $|\mathbf{R}_{ground}|$ 的增加，性能的提高降低，更少的利用 X 和 Y 之间的交互信息。

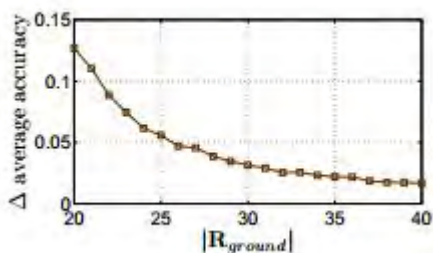


图 3: 如何在提高精度 DAKM (在 X 和 Y 的平均值) 与 X 和 Y 之间的关系的数量变化

接下来我们考虑 DAKM 是如何受个体聚类任务的困难度影响的。我们生成 4 个视图，基于他们对聚类准确是多么的困难：硬 (H) X 和 Y，一个简单的 (E) X 和 Y 容易。硬聚类的任务是在这样一种方式，聚类比简单的工作更易产生较大的重叠。我们展示这些不同的组合视图聚类结果见表 1，精度显示为 k-均值和 DAKM，我们还展示了两者的差异。可以看出，DAKM 在所有情况下显著改善普通 k-均值。这表明 DAKM 鲁棒性：人们所期望的，例如，从一个更困难的工作的噪声会降低一个容易的工作绩效，但这显然不是这样。

Task Diff.	k-means		DAKM		Δ	
	x	y	x	y	x	y
H x, H y	56.7	60.0	59.3	62.4	2.7	2.4
H x, E y	56.5	92.8	63.3	95.9	6.7	3.2
E x, E y	83.0	92.8	90.9	97.5	7.9	4.7

表 1: 当视图变难，DAKM 的效果改变。

4.2 人的行为聚类

4.2.1 数据集

我们使用三个真实的数据集实验。UCF YouTube [5], UCF Sports [11], and Hollywood-2[7]。我们的报告只针对行动聚类的准确性，而现场地面的准确性是不提供的。UCF YouTube 数据集包括 1168 个各种人体活动的视频。UCF 体育数据集包括 150 个各种体育视频，从各种广播源。Hollywood-2 数据集包括从好莱坞电影里收集的 1707 个动作视频。

4.2.2 安装程序

从一个动作中提取一个动作表示，我们使用公开的代码于默认的设置提取密集的轨迹特征如 Wang et al. [12]所展示的那样。然后我们处理一个词袋式表示的特征。我们进行主成分分析后 k-均值聚类的每个描述符分别(HOG, HOF, MBH, Tr)，我们捕捉 95%的 PCA 和 K = 2000 方差聚类。这个表示在一个 8000 维 (4x2000) 的运动频率直方图特征。现场展示中，我们使用 SIFT 特征【6】，在 10 帧的时间间隔从每个视频提取。再次实验，进行 PCA 和 k-均值，k = 2000，显示在一个 2000 维的频率直方图的静态特征，我们正常化直方图的运动和静态特性。

得到的频谱表示，我们首先找到成对的距离之间的所有直方图，利用直方图相交的一组：

$$S(a, b) = \sum_{i=1}^n \min(a_i, b_i) \quad (9)$$

我们应用公式 9 来获得一个行动和场景的相似图形。然后我们进行 Ng et al. [8]的步骤 2-4，在动作和场景上得到两个相互地里的光谱表示。然后我们在得到的载体上进行 DAKM，我们设置 k_a （行动聚类的数目）为数据集上动作类型的数目。地面标定没有提供数据集的场景，我们没有 k_s 的先验知识，但初步的实验表明，一个相对高一些的场景是最好的-我们设置所有的数据 $k_s = 40$ 。所有的算法是随机的，所有的实验都运行 10 次，结果取平均值。

4.2.3 人类行为的聚类分析结果

在所有三个数据集聚类的运动结果总结在表 2。我们比较以下的聚类方法：谱单视图 (SSV) Ng et al. [8]提供的运动特征；光谱级联的视图 (SCV) Ng et al. [8]提供的运动和静止的直方图。共培训光谱多视图 (CMV) 最近提出的多视图聚类方法 Kumar and Daume III [4]，运动和静态特征作为动作的两种视图；SDAKM $\lambda = 1000$ (SD1k)；SDAKM λ 设置为每个数据集最优值(SDO)。SDO，我们考虑以下值 λ ：1, 5, 10, 50, 100, 500, 1000, 5000, 10000 和 50000。

Dataset	Clustering Accuracy (%)				
	SSV	SCV	CMV	SD1k	SDO
YouTube	39.2	40.7	38.1	41.8	43.9
UCF Sports	68.0	67.2	64.2	72.9	76.0
Hollywood-2	35.6	32.4	31.5	36.5	36.5

图 2：现实的数据集的各种方法的聚类性能

从表中可以看出，SDAKM 最佳 λ (SDO) 给出了在所有三个数据集最高的精度，我们提出这是因为它最能有效利用动作和场景之间的复杂关系。在数据集上 CCM 性能院差于基

线聚类方法 (SSV) 显示出多角度的假设不适用于动、静特性之间的关系。链接运动和静止矢量 (SCV) 也会对两个或三个的数据集产生负面影响。SDAKM 在 UCF 体育数据集上给出基线的最大提高, 而数据集的两个属性: 高度稀疏的关系 R_{ground} 和简单集群的场景。另外, 最弱的性能数据上看到观察 hollywood-2, 只增加了 0.9% 的精度, 即使有最佳的 λ , 这不足为奇: Marszalek et al. [7] 使用场景来提高 Hollywood-2 数据集的精确度, 利用标定数据直接推断动作和场景之间的关系但只观察了基线性能提高 1.1%。

在图 4 我们展示不同的参数 λ 对不同数据集 SDAKM 性能的影响。峰值性能在每个数据集是不同的, 这是预期的观察, 未知 R_{ground} 的稀疏度在数据集之间可能是不同的。在今后的工作中, 有利于自动使用两视图的数据集之间的交互信息估计最优的 λ 值。然而, 当 $\lambda \geq 1000$ 时, SSV 对数据集仍有改进的意义。

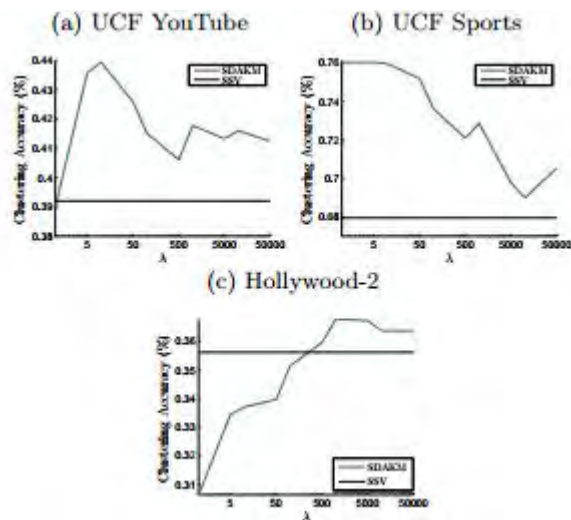


图 4: 各种 λ 下 SDAKM 的性能

为了解 λ 进一步的影响, 我们考虑 $\lambda = 1$ 时 YouTube 数据集的聚类, 我们建议在这种情况下, R' 趋于比行动和场景之间的真实关系更稀疏, 这将扭曲聚类结果。图 5 说明了聚类和 $\lambda = 1$ 的 YouTube 数据集的迭代的性能。继续迭代 R' 更加稀疏, 最初, 这个结果更准确的聚类 (3.6% 比初始解第三次迭代后), 当算法终止, 它已经下降到 1.7% 低于初始精度, 表明 R' 不再反映场景和动作之间的真实关系。

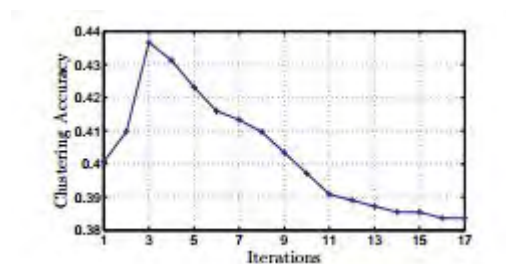


图 5: $\lambda = 1$ 几个迭代后峰值性能 YouTube 数据集 SDAKM, 但算法终止性能差。

虽然数据集没有地面真实的场景在图 6 中，我们提供一些在 UCF YouTube 数据集发现场景类别的例子。几个关键场景类别清晰起来：镜头包括天空；庭院中玩耍；田地；泳池；蹦床。这些都多多少少与数据集中的某些行为有关，正如人们所预料的。例如，天空场景通常显示高尔夫，网球或足球杂耍，但球场的场景通常表明排球，网球，篮球。此外，游泳池和蹦床的场景有潜水和跳跃分别接近一一对应。显然在这种情况下额外的现场信息可能有助于聚类行为。



图 6：场景类别在 UCF YouTube 数据集发现的例子，和他们最密切相关的行动。

5. 论述

本文提出了一种新的算法已被提出双分配 k -均值聚类，根据一个数据集的两个视图产生的两个聚类方案。不同于以往的方法，这是用于当一个数据集有两个共同的视图，每个视图有一个单独的聚类方案-类似其他聚类方法只适用于单一视图生成多个聚类解决方案（替代聚类）或产生一个单一的聚类的多个视图（多视图聚类）。我们已经证明 DAKM 对合成数据和真实的人类动作场景的数据集可以显著的提高聚类结果。这种性能的改善是明显的即使在双方视图的聚类分离较差，演示 DAKM / SDAKM 鲁棒性。我们进一步的工作将集中在确定 λ 的自动，我们相信这可以计算数据集的两个视图之间的交互信息。同时，在这里提出的算法的限制，对于复杂的原因，考虑双分配，今后我们计划要考虑多个分配聚类。

工具书类

- [1] 美国 Bickel 和 T. 希尔。多视图聚类。在诠释中。数据挖掘，页 19-26，2004 **2**
- [2] H. 党和 J. 贝利。使用 CAMI 方法聚类方案生成。在暹罗国际会议的数据挖掘，页 118-129，2010。 **2, 4**
- [3] N. Ikingler-Cinbis 和 S. Sclaro 对象，场景和动作：结合人体动作识别的多特征。在过程。欧洲会议计算机。视觉，页面 494-507，2010。 **2**
- [4] A. Kumar 和 H. Daume III. 多视图的谱聚类的协同训练方法。在过程。马赫诠释中。学习，页 393-400，2011。 **2, 6**
- [5] J. Liu, J. Luo, 和 M. Shah. 识别的实际行动视频在野外”。在过程。IEEE 会议计算机。视觉与模式识别，页 1996 - 2003，2009 六月。 **6**
- [6] D. G. Lowe 独特的图像特征从 scaleinvariant 要点。Int. J. 计算机视觉。60:91-110, 2004。 **6**
- [7] M. Marszalek, I. Laptev, 和 C. Schmid. 活动背景。PROC. IEEE 会议计算机。视觉与模式识别，（我）：2929-2936，六月 2009。 **1, 2, 6, 7**

- 【8】A. Y. Ng, M. I. Jordan, 和 Y. Weiss. 谱聚类算法和分析。在提出神经网络。过程。系统, 页面 849-856, 2001。5,6
- [9] J. C. Niebles, H. Wang, 和 L. Fei-Fei. 用时空的话, 人类行为的类别的无监督学习。Int. J. 计算机。视觉, 79 (3) : 3 299-318, 2008。2, 5
- [10] A. Prest, C. Schmid, 和 V. Ferrari 弱监督人和物体之间的互动学习。电气和电子工程师协会。模式分析。马赫。智能。 , 34 (3) : 601-614, 2012。2
- [11] M. D. Rodriguez, J. Ahmed, 和 M. Shah. 行动马赫: 时空最大平均相关高度滤波器的动作识别。在过程。IEEE 会议计算机。视觉与模式识别技术, 2008.6
- [12] H. Wang, A. Klaser, C. Schmid, 和 C.-L. Liu. 密集的轨迹动作识别。在过程。IEEE 会议计算机。视觉与模式识别, 页面 3169-3176, 2011。6
- [13] Y. Wang, H. Jiang, M. S. Drew, Z.-N. Li, 和 G. Mori 动作类的无监督的发现。在计算 proc.ieee 会议。视觉与模式识别, 页面 1654-1661, 2006。2
- [14] Y. Yang, I. Saleemi, 和 M. Shah 无监督的分组和一次人类行为, 手势学习发现运动原语, 和表情, IEEE 会议, 模式分析。马赫。智能。 , 35 (7) : 1635-1648, 2013。2, 5