一种视频数据代表选择框架方法

蒋 勇[1] 张海涛[2]
(西南政法大学 重庆 401120)[1] (中国人民公安大学 北京 100038)[2]

摘 要 为有效处理视频数据问题,提出一种识别海量数据集中代表子集的方法,即代表选择方法,经选择后的小容量的数据代表完全可以代表原始大数据集的结构特征。对于给定的大数据集,首先生成相应 ℓ_1-norm 非负稀疏图,然后利用一种谱聚类算法基于所生成的稀疏图将大数据反复划分直至形成聚类簇。代表选择过程中,将每个聚类看作 Grassmann 流形中的一个点,然后基于测地距衡量这些点间的距离,接着利用 min-max 算法分析距离以提取出较优的聚类子集。最后,通过分析被选集类的一个稀疏子图,利用主成分集中性方法探测出数据代表,称此过程为基于非负稀疏图与 Grassmann 流形测地距的代表选择框架。为验证所提出的框架,将之应用于视频分析中,从一长段的视频流中识别出少数的几个关键帧,实验效果通过人工判断与标准评价方法进行评价,并与现有的几种方法的效果进行比对,结果证明所提出的代表选择框架方法具有更好的效果与可行性。

关键词 稀疏图,Grassmann 流形,测地距,关键帧,代表
中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.11.004

Representative Selection Framework Approach for Videos

JIANG Yong[1] ZHANG Hai-tao[2]
(Southwest University of Political Science and Law, Chongqing 401120, China)[1]
(People's Public Security University of China, Beijing 100038, China)[2]

Abstract In order to solve the process problem of massive videos data, a representative selection method of identifying the optimal subset of data points as a representative of original massive dataset was proposed. The selected data points of subset can represent inner structure of original massive dataset. And the novel representative selection method is based on ℓ_1-norm non-negative sparse graph for the original massive dataset. The massive data points are partitioned into some clusters by using a spectral clustering algorithm based on the non-negative sparse graph generated in previous steps. Each cluster is viewed as a point in the Grassmann manifold, and the geodesic distances among these points are measured. By using a min-max algorithm, geodesic distances are analyzed to build an optimal subset of clusters. Finally, the principal component centrality method is used to detect a representative after analyzing the sparse graph of selected clusters. The proposed framework is validated on the problem of video summarization, where a few key frames should be selected in long video clips which contain massive frames. The comparison of the results obtained between the proposed algorithm and some state-of-the-art methods was producted. Result indicates the effectiveness and feasibility of the proposed framework.

Keywords Sparse graph, Grassmann manifold, Geodesic distance, Key frame, Representative

1 引言

从大量视频数据中获取、存储、提取有价值的信息是一项具有挑战性的工作,关于此类数据的处理通常涉及3个方面的问题,即容量、速度与多样性。大容量、高维性显而易见地成为视频数据的基本特性。而数据的多样性是指数据来源广泛,从传统类型的文本业务数据到多媒体如语音、视频与图像数据等[1]。

视频数据往往较难处理,大量有价值的信息并没有被提取出来。因此,如何从视频数据集中选取一个子集,且这个子集作为整个视频数据源的代表模型,能够刻画原数据的结构特点,是视频数据集处理研究的一个重要方向。

目前,有一些研究工作试图利用数据代表模型抽取来解决视频数据集的应用问题[2],但这些方法需要创建一个数据点的密度相似性矩阵,由于资源消耗大,限制了数据集应用规模。尤其是近年来多媒体数据(图像、音频和视频)飞速增长,体量巨大且多种多样,相比模式单一的结构化数据包含更丰富的信息,对此类数据进行分析以期挖掘出有价值的信息面临着巨大挑战。多媒体分析研究包含的种类很多,如多媒体摘要、多媒体索引和视频检索、事件检测等。其中,视频摘要

到稿日期:2016-05-07 返修日期:2016-07-31 本文受重庆市教委科技项目:基于大数据的职务犯罪情报分析模型与供给式研究(KJ1600103),公安部公安理论及软科学研究重点项目(2013LLYJGADX003)资助。

蒋 勇(1986—),男,博士,讲师,主要研究方向为情报分析、电子物证,E-mail:xizheng23@163.com;张海涛(1982—),男,博士生,主要研究方向为信息安全、数据挖掘。

任务最具有代表性，其可以理解为提取原始序列中最重要或更具代表性的静态或动态内容。海量视频数据集明显符合"3Vs"模式的大数据标准，其分析处理过程对于一般聚类或数据挖掘方法难以奏效。因为是取数据全体而不是取样，算法高效才是视频数据集处理追求的目标。而面对大规模视频数据集分析问题，一般方法通常关注于关键帧的提取，如文献[3]中提出的稀疏代表方法(Sparse Representation Based Method，SR)从视频数据中选择包含信息最大的帧子集，具体过程是从无结构的用户视频中，基于所提出的稀疏代表方法来提取关键帧；在实现过程中，视频帧被映射到一个低维随机特征空间，将稀疏信息代表理论用于分析视频数据的时空信息并生成关键帧，该方法具有较好的计算效率，不需要镜头探测、分割或语义理解等处理工作。

用户视频的质量被视作很多因素共同作用的结果，如摄像机移动、场景干扰、移动对象交互、图像质量、镜头设置等。因此 Luo 等人[4]提出一种基于动作的关键帧提取算法(Motion based Key Frame Extraction，MKFE)为一组类型的用户视频构建了基准对象。通过评估摄像头动作类型，如跨度、聚集、停顿与固定等动作，视频片段被分为几个不同类型。然后，从这几个不同性质的类型中分别提取关键帧，使得关键帧成为捕捉重要与新类型事件的最有价值的信息帧。在一段视频序列中，靠近开始与结尾的几个帧可能在关键帧选择中发挥关键性作用。这是本文方法所贯穿的思想。实际操作中，评判者或鉴定者往往会在序列的开头与结尾处选择关键帧。对于每个视频片段，评判者同时也需要将帧选择任务与视频摘要问题相结合以作出决定。

文献[5]提出一种两层组稀疏代表方法(Bi-Layer Group Sparsity，BGS)基于可视内容从视频中提取关键帧。这种两层组稀疏代表方法是基于传统的 Moreau-Yosida 正则化组稀疏 LASSO 方法，将输入视频中时空关联关系模型化以提取关键帧。此方法中，输入视频中帧首先被分成多个具有同类性质的块，并同时将组稀疏以两层方式进行：块到帧，从帧到序列。分组的稀疏系数会进一步结合帧语义质量得分情况来提取关键帧。该方法不需要依赖任何非可视数据或者领域知识，且比现有方法更具健壮性。

文献[6]提出的稀疏模型代表选择方法(Sparse Modeling Finding Representatives，SMFR)将关键帧提取问题看作从数据集中选择少数代表问题的一种特例，这种数据集可以是视频序列，也可以是图像或一般数据集。考虑为一个数据集寻找一些代表的问题，且要求一个数据点子集可以有效地描述整个数据集，可假定每个数据点被表达为这些代表的线性组合，并且将代表选择问题形式化为一种稀疏多向量计算问题。在形式化处理中，字典与量度都在数据矩阵中给出，且未知的稀疏编码通过凸优化方式选择代表。一般不假定数据是低秩的或者是围绕聚类中心分布的，但当数据确实来源于低秩模型集合时，该方法可以从每个低秩模型中自动地选择一些代表，并分析这些代表的几何结构，讨论它们与凸面顶点的关联关系。实验证明了所提框架可以有效地探测处理数据集中的异常点，并可泛化到大型数据集中。基于提出的框架与理论方法，所提取的代表可以有效地用于视频摘要与图像分类处理问题中。

本文提出一种数据模型选择方法用于处理一般大数据问题，其具有一定通用性。其中，基于 ℓ_1-norm 非负稀疏图[7]设计一种谱聚类技术来处理数据集，其性能相比传统图方法的性能更好。此方法中，每个聚类都被看作 Grassmann 流形中的一个点，用来衡量聚类点之间的测地距。并利用 min-max 算法[8]结合测地距来检出聚类的代表模型子集，每个被选定的聚类都与原始稀疏图的一个子图相关联；接着，子图主成分集中性方法(Principal Component Centrality，PCC)被用来选择稀疏子图的代表模型。第 2 节图 1 描述了提出的利用非负稀疏图与 Grassmann 流形测地距解决代表选择问题的方案框架。方法的验证思路是基于具体的大数据应用问题来衡量系统的综合性能，即将所提出的模型框架方案应用于大型视频数据分析中，解决在大量视频序列中选择一组关键帧频的问题。实验结果由人工判断与标准评价方法进行评价，且将本方案与现有方法进行比较，证明了所提出方案的可行性。

本文第 2 节提出基于 ℓ_1-norm 非负稀疏图、Grassmann 流形测地距及 PCC 的模型选择框架；第 3 节基于实验将多种现有方法进行分析比较；最后总结全文并展望未来工作。

2 数据代表选择框架

文中提出的大数据代表选择框架如图 1 所示，原始数据经规范化方法获取非负稀疏表示，然后基于谱聚类方法形成初始聚类集合，最后利用测地距方法与 PCC 方法产生最终的数据代表。具体方法从 2.1 节开始进行进一步阐述。

图 1 数据代表选择框架

2.1 基于 ℓ_1-norm 非负稀疏图的数据聚类

聚类技术是大数据分析的主要手段之一。模型代表的选择问题与聚类技术有着较强的关联性，其中每个代表都可以看成一个聚类体，以前的一些方法主要通过衡量数据集中点对相似性来创建一个相似性矩阵进行计算，但对于大的密度矩阵而言，其耗费计算资源太多，特别是对于大数据来说更难以实现。本节利用近期在图像分析中成功运用的数据的自我表达思想及稀疏子空间聚类来创建一个稀疏关联矩阵，以有效解决数据聚类问题[9]。

影响聚类质量的潜在因素是如何定义数据领域，直接的方法是利用点对欧几里德距离，即一个点与其他点的基于 k 最近邻计算或 ε-球基距离计算方法，前者是用 k 个最近邻点与中心点的距离，后者是指样本在 ε-球半径范围内邻接点的比较，此类方法受限于一个预先定义邻域的节点个数，或者一个固定半径阈值的设置，由此，这些方法并不适合内部数据邻接结构多样化的大数据集。这里利用一些基本方法创建一个非负稀疏图试图解决问题，过程中不需要进行矩阵的逆运算，可以避免矩阵的奇异值与小样本问题[10,11]。

用 $U=\{h_j\in\mathbb{R}^D\}_{j=1}^N(j=1,2,\cdots,n)$ 表示数据集，其中，D 为欧几里德空间，其比数据集中元素数 N 小，即，$N\gg D$，为每个点定义一个领域的出发点是考虑了数据本身作为稀疏表示的字典。数据集可以用矩阵 $H(H=[h_1,h_2,\cdots,h_N]\in\mathbb{R}^{D\times N})$ 表示，令 $H_i^{\wedge}\in\mathbb{R}^{D\times(N-1)}=H/\{h_i\}$ 为从 H 原矩阵中去掉第 i 列所得，算法试图通过其相应字典 $H_i^{\wedge}$ 为 h_i 寻找其最稀疏表示模型：

$$\min \| c_i \|_0 \text{ subject to } h_i = H_i^{\wedge} c_i \tag{1}$$

其中，$\| \cdot \|_0$ 为 ℓ_0 范式，用于计算零元素个数，尽管这个问题为 NP 难题，但通过一些比较研究发现，最接近解决方案可以通过 ℓ_1 范式最小化值得到：

$$\min \| c_i \|_1 \text{ subject to } h_i = H_i^{\wedge} c_i \tag{2}$$

ℓ_1 范式的最小化可等价性转化为松散形式的凸优化问题，对于一个固定字典 $H_i^{\wedge}$，有：

$$c_i = \arg \min_{c_i \in \mathbb{R}^{N-1}} \| h_i - H_i^{\wedge} c_i \|_2 + \lambda \| c_i \|_1 \tag{3}$$

这里存在一个全局优化解决方案，采用一个有效的 ℓ_1 范式优化工具箱，基于一个数据集可以将创建一个非负稀疏图。过程总结如下：

1)输入数据集以矩阵形式表示，$H=[h_1, h_2, \cdots, h_N] \in \mathbb{R}^{D \times N}$。

2)对于每个数据点 h_i，式(3)为其相应非负系数 $c_i \in \mathbb{R}^{N-1}$，其在系数矩阵 $C \in \mathbb{R}^{N \times N}$ 中第 i 列按序排列，其中在 C_i 的第 i 个位置插入 0 值，也就是说矩阵 C 是 0 对角矩阵。

3)以 $G=\{H, \tilde{C}\}$ 的形式构建图，其中 H 中的每个点都被映射成一个顶点，且 $\tilde{C}=[\tilde{C}_{ij}]_{N \times N}$ 表示图中权重矩阵，且有 $\tilde{C}_{ij} = |C_{ij}| + |C_{ji}|$。

本文利用一种谱图理论算法进行数据分割，对于子图 A，B 和顶点 V，有：

$$Ncut(A,B) = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(A,B)}{assoc(B,V)} \tag{4}$$

其中，$assoc(A,V) = \sum_{u \in A, t \in V} c(u,t)$。

规范割集准则不仅可以衡量类内样本间的相似程度，同时也可计算类间样本的相异程度[12]。特别地，归范割算法反复地将输入数据集分成两个聚类，并且检出聚类中元素分布线性空间的最大秩，如果其中一个聚类的秩超过了阈值，则此聚类将被迭代划分成较小的聚类簇，这个过程是为了避免一个聚类中心含有过多的数据点。接下来介绍一个算法用于选择较重要的聚类中心，然后从这些聚类簇中选择模型数据代表。

2.2 利用 Grassmann 流形测地距方法选择最优聚类中心子集

Grassmann 流形 $Grass(p,n)$ 为空间 $\mathbb{R}^n$ 中的 p 维线性子空间的一个集合，流形上的每个数据点代表了 $n \times p$ 维的正交矩阵的列向量张成的子空间[13]。本节将每个聚类作为 Grassmann 流形中的一个点，形成的聚类的数量比最终所需代表数要多，有些聚类包含了冗余信息，所以要缩简。本文利用 Grassmann 流形测地距来衡量两个聚类间的相异性，然后基于 min-max 算法得到最终优化的聚类子集。流形 $Grass(p,n)$ 中的元素 $\tilde{y}$，即一个 p 维 $\mathbb{R}^n$ 子空间，可以被一组 p 个向量 $y_1, \cdots, y_p$ 确定，而 $\tilde{y}$ 可以是它们的线性组合。这些向量可作为一个 $n \times p$ 维矩阵 Y 的列向量，而 $\tilde{y}$ 可称为 Y 的列空间或者跨度，可写为 $\tilde{y} = \text{span}(Y)$。如果 Y 是满秩，则矩阵 Y 的跨度称为 $Grass(p,n)$ 流形的一个元素，这些矩阵组合被定义为一种非紧凑型的 Stiefel 流形：

$$ST(p,n) := \{Y \in \mathbb{R}^{n \times p} : rank(Y) = p\} \tag{5}$$

给定 $\tilde{y} \in Grass(p,n)$，矩阵 Y 位于流形 $ST(p,n)$，流形中矩阵集合具有与 Y 同样的跨度：

$$YGL_p := \{YM : M \in GL_p\} \tag{6}$$

其中，GL_p 表示 $p \times p$ 可逆矩阵集合，这将流形 $Grass(p,n)$ 等价于商空间 $ST(p,n)/GL_p := \{YGL_p : Y \in ST(p,n)\}$。四元组 $(GL_p, ST(p,n), \pi, Grass(p,n))$ 被称为一个主 GL_p 分量，总空间为 $ST(p,n)$，基空间 $Grass(p,n) = ST(p,n)/GL_p$，组动作有：$ST(p,n) \times GL_p \ni (Y,M) \mapsto YM \in ST(p,n)$ 且有映射：$\pi: ST(p,n) \ni Y \mapsto span(Y) \in Grass(p,n)$。

令 $W \in ST(p,n)$，矩阵 W 定义一个仿射交叉面：

$$S_W := \{Y \in ST(p,n) : W^{\mathrm{T}}(Y - W) = 0\} \tag{7}$$

与 WGL_p 正交，令 $Y \in ST(p,n)$。如果 $W^{\mathrm{T}}Y$ 可逆，则等价类 YGL_p（与 Y 具有同样跨度的矩阵集合）与交叉面 S_W 在单点 $Y(W^{\mathrm{T}}Y)^{-1}W^{\mathrm{T}}W$ 处相交。如果 $W^{\mathrm{T}}Y$ 是非可逆的，则位于 WGL_p 与截面 S_W 之间的交集为空，令 $\tilde{u}_W := \{span(Y) : \exists W^{\mathrm{T}}Y^{-1}\}$ 代表 YGL_p 与截面 S_W 的交叉子空间集合，映射如下：

$$\sigma_W : \tilde{u}_W \ni span(Y) \to Y(W^{\mathrm{T}}Y)^{-1}W^{\mathrm{T}}W \in S_W \tag{8}$$

此映射称为交叉面映射，实现了 $Grass(p,n)$ 流形子集 $\tilde{u}_W$ 与 $ST(p,n)$ 仿射子空间 S_W 之间的一一对应映射；对于 $W \in ST(p,n)$，经典的 $Grass(p,n)$ 流形结构令 σ_W 成为 $\tilde{u}_W$ 与 S_W 间的微分同胚。$Grass(p,n)$ 的参数化由下式给出：

$$\mathbb{R}^{(n-p) \times p} \ni K \mapsto \pi(W + W_{\perp}K) = span(W + W_{\perp}K) \in \tilde{u}_W \tag{9}$$

其中，$W_{\perp}$ 为 $ST(n-p,n)$ 中任一元素，且有 $W^{\mathrm{T}}W_{\perp} = 0$。

这里，数据的近似切空间采用主成分分析法来获取，且数据点之间的关系转换成子空间之间的关系，在局部范围内这种度量具有唯一性。在近似切空间中，令每一点处切空间基向量数量相同，则每一点处的切空间可看作 Grassmann 流形 $G_{n,m}$ 上的一点，数据点间的相似度计算通过测量 Grassmann 流形上点之间的距离来实现，本文将两个子空间之间的距离映射为流形中两点间测地距，后者主要通过主角定义，主角由两个子空间的点积的奇异值分解计算得到。

定义两个子空间 H_1，H_2，此处假定，$\dim(H_1) = d_1 \geqslant \dim(H_2) = d_2$ 且 $(u_1, u_2, \cdots, u_{d_2})$，$(v_1, v_2, \cdots, v_{d_2})$ 分别是子空间 H_1 与 H_2 的正交基向量，两个子空间 H_1 和 H_2 间的主角 $0 \leqslant \theta_1 \leqslant \cdots \leqslant \theta_t \leqslant \pi/2$，有：

$$\cos\theta_t = \max_{u_t \in H_1} \max_{v_t \in H_2} u_t^{\mathrm{T}} v_t \tag{10}$$

$$\text{s.t. } \| u_t \|_2 = 1, \| v_t \|_2 = 1, u_j^{\mathrm{T}} u_t = 0, v_j^{\mathrm{T}} v_t = 0 \text{ for } j = 1, 2, \cdots, t-1$$

主角的映射度量依据以下公式计算：

$$d = \sqrt{\sum_{i=1}^{d_2} \sin^2\theta_i} = \sqrt{d_2 - \sum_{i=1}^{d_2} \cos^2\theta_i} \tag{11}$$

基于上述映射，本文定义 Grassmann 流形测地距如下：

$$G(H_1, H_2) = \sqrt{\frac{1}{d_2} \sum_{i=1}^{d_2} \cos^2\theta_i} \tag{12}$$

测地距在先前图像处理中已经成功运用，用来处理数据划分树中的叶节点。作为一种度量方法，它具有一些良好的特性，如对称性、三角计算，它来源于 Grassmann 流形固有的几何特性，即代表了连接流形上两点之间测地线的距离，保持了数据内在的非线性几何结构。

本文结合上述 min-max 算法，利用测地距来选择优化的聚类子集。不同于文献[8]中所述方法，这里的相似矩阵是基于聚类集合两两距离计算而得，显然，这种方式比直接求点集的计算所消耗的计算资源要小得多。算法 1 总结了优化的聚类子集选取过程。

算法 1 min-max 算法的最优子集选择过程

输入：聚类集合(Grassmann 流形数据点)，结果聚类数

输出：最终聚类子集

Begin：

1. 基于测地距创建相似矩阵
2. 检测出具有最大测地距度量的两个聚类
3. 重复过程，直到所有聚类簇都筛选过
4. 扫描所有剩余聚类
5. 选择其中一个聚类，其满足与先前所选聚类集的距离最大化

End

2.3 主分量中心性用于代表选择

最后一步是从聚类簇中选择数据代表，这里，每个聚类在先前步骤中都被映射到 ℓ_1-norm 的非负稀疏子图中，然后基于主分量中心性原则评估顶点位置的重要性[14]。主分量中心性来源于信号领域的卡·洛变换，PCC 采用的是网络的结构信息，提供一个通用框架将图转化为一个谱空间，以便分析图的中心性，$A\in\mathbb{R}^{m\times m}$ 表示稀疏子图 $\tilde{C}$ 的邻接矩阵，设 A 的特征值 $|\lambda_1|\geqslant|\lambda_2|\geqslant\cdots\geqslant|\lambda_m|$ 分别对应于归一化的特征向量 $x_1,x_2,\cdots,x_m$。并令 $\Lambda=[\lambda_1\lambda_2\cdots\lambda_m]^{\mathrm{T}}$ 表示特征值向量，$X_{m\times q}$ 表示 X 的子矩阵，包含 X 前 m 行前 q 列。然后，PCC 可以用矩阵形式表达如下：

$$C_q=\sqrt{((AX_{m\times q})\cdot(AX_{m\times q}))1_{q\times 1}} \tag{13}$$

或者

$$C_q=\sqrt{(X_{m\times q}\cdot X_{m\times q})(\Lambda_{q\times 1}\cdot\Lambda_{q\times 1})} \tag{14}$$

算符“·”为哈达马得(元素对应相乘)算符，图中一个节点的 PCC 值为该节点从 q 个最重要特征向量构成的 q 维特征空间原点间的欧几里德距离。

2.4 框架方法分析

如果根据谱图理论直接对海量数据进行特征值和特征向量处理，将耗费不可估量的计算资源，实际上变得越来越不可行，故研究大规模数据学习与处理的高效、可扩展的方法成为现实需求。基于数据自我表达思想的模型创建方法资源耗费小、效率高，有较好灵活性，可以用来解决大数据集问题。

本文首先通过数据的稀疏图表达，割集的聚类划分而渐进性地将处理对象转化为数据中心簇的低维形式，即使到这一步，聚类簇数目仍然庞大，而流形学习的作用即是发现海量高维数据集中的整体几何架构和分布规律，其目标是将高维数据空间中低维度结构呈现出来，基于局部的方法可进一步降低聚类簇的计算复杂性，并使中心数据间的相似度计算转化为流形中点距计算来实现，维持了数据内在结构特性。

所提出的框架方案中最后一步是基于主分量中心性计算评估聚类中心的重要性，将最有代表性的子中心选择出来，最终形成目标数据代表。虽然涉及到特征值处理，但经过前期的分解提取，此时的计算已经不再消耗大量资源，复杂度已大大降低。给出的框架方法可最大限度地减少直接处理大规模数据的难度，充分利用了数据间的自我表达特点，减小维度的同时保留了数据拓扑结构，通过一种抽丝剥茧式的提取过程，使得大规模问题转化成小样本选择处理过程。

3 实验分析

本文将所提出的框架方法用于视频摘要问题以验证方法的有效性。视频摘要问题是视频相关应用中一个广泛存在的应用领域。本文尝试将提出的代表选择方法用于视频摘要中。从大数据角度出发，目前视频摘要研究中存在一个内在的局限性，即不能将之泛化至代表选择问题中，首先，大多数视频摘要技术都具有领域依赖性[15,16]，因为利用了相关领域视频片段的特点；其次，即使这些技术产生了可接受的效果，其付出的计算复杂度也是可观的，有时创建一个视频摘要的时间可达到视频长度的 10 倍以上。

3.1 视频数据库

实验所用视频资料来源于开放视频项目共享数字视频库[17]，选择其中 7 个视频段，每个时间长度为 2min，相关细节如表 1 所列，视频片段是 MPEG 压缩格式。

表 1 实验所用视频集

视频片段	视频文件	持续时间 mm:ss	分辨率	帧总数	过渡域数量
Exotic Terrane, Segment1	UGS01-001	02:40	320 * 240	4798	5
Exotic Terrane, Segment2	UGS01-002	02:25	320 * 240	4435	5
Exotic Terrane, Segment3	UGS01-003	03:10	320 * 240	5695	5
America's New Frontier, Segment1	UGS02-001	02:05	352 * 240	3686	3
America's New Frontier, Segment2	UGS02-002	02:41	352 * 240	4830	3
America's New Frontier, Segment3	UGS02-003	03:49	352 * 240	6879	3
America's New Frontier, Segment4	UGS02-004	04:13	352 * 240	7608	3

这些视频片段从 3686 帧到 7608 帧不等，平均长度为 5418 帧，这里不预设每段关键帧数量。区别于大多数方法，所提方法不需要事先对数据预取样，可直接将之扩展应用于规模更大的视频片段。给定输入视频序列，从每帧中提取特征向量以减少像素域的过高的维数，尤其选择着色直方图作为特征向量，为每个 RGB 组件使用 16 个描述符，然后将每个帧映射到位于 R^{48} 欧几里德空间的一个向量点。在视频库中，瞬时域内的两个帧的特征非常相近，因此，对于每个帧，它的邻接范围预定数量的帧将从相应稀疏代表字典中移除。实验中，对于每个帧，它的邻域中至少 15 个连续帧将从字典中移除，另外，每个稀疏因子将根据时间下标而定，这里，$C_{ij}:=e^{\beta|i-j|^2}C_{ij}$，其中将 $\beta=0.02$ 作为一个常量而设定。

在谱聚类步骤中，将上限秩阈值设为 10，反复利用规范化切算法。上限秩控制了每个聚类中最大元素数，因此可以自动决定最终的聚类数目。本文也评价了通过选择一个预定聚类数来基于上限秩值进行迭代数据聚类所带来的影响，利用一个预定义的聚类数目，反复对输入视频序列进行划分，此过程的效果较好。

3.2 评价方法

基线算法：与现有的相关方法进行比较，包括 MKFE，SMFR，SR 及 BGS 方法。对于大规模数据的代表提取问题而言，一直缺乏客观的评价标准，大多采用专家判断的方法，评价过程中，通过将一致认定的基准关键帧与算法生成的关键帧做比对，对其做出“好”、“中”、“坏”或者 1，0.5，0 等评分。这些评分之后被用于大规模数据集中的算法效能评估，尽管

这也许是最实际的评价方法，特别是基于用户角度提取的关键帧技术，但这种评价方法却难以用于更广泛的视频应用研究中，原因是实验参数过多地掺杂了人工干预因素。本文构建自己的评价标准。

保真度：用于衡量视频的全面覆盖率，基于半豪斯多夫距离算法[18,19]，来计算关键帧集与原始视频帧集之间的相似程度（最大距离），即有：

$$d_f=\max_i(\min_j(d_{ij})),\forall i\in[1,N],\forall j\in[1,m]$$

其中，$d_{ij}=D(f_i,Kf_j)$是指第 κ 聚类中第 i 帧与所提取的第 j 关键帧之间的非相似度度量。一般地，保真度用下面的规范化格式计算：

$$Fidelity=\frac{1}{K}\sum_{i=1}^{\kappa}(1-\frac{d_f}{\max_i(\max_j(d_{ij}))}) \tag{15}$$

较高的保真度说明所提取的关键帧能够对视频序列的内容提供更好的全局性描述。

压缩率：指视频数据代表对原始大数据的描述简洁度，计算如下：

$$CR=\frac{m}{N} \tag{16}$$

其中，m 指算法检出的正确关键帧数量，数量相对原始视频序列极少，故可以进行人工干预；N 指原始视频中帧的数目，对海量数据而言，N 可以为估计常量。

3.3 实验结果

计算复杂度：因产生一系列关键帧所需时间依赖于特定的硬件，故在关于计算复杂度方面绝对公平地评价各个不同方法的优劣几乎是不可能的。本文通过对每帧的平均处理时间来评价算法复杂度。

在实验中，所提方法对于每帧的处理时间平均为0.0140s，表2为帧平均处理时间比对结果，实验平台统一为 Intel Core E7500@2.93GHz，平均处理时间还可以通过预取样工作而缩短。

表2 视频序列帧的平均处理时间

视频文件	MKFE	SMFR	BGS	SR	新方法
UGS01-001	0.0201	0.0212	0.0156	0.0167	0.0153
UGS01-002	0.0172	0.0146	0.0152	0.0148	0.0131
UGS01-003	0.0153	0.0137	0.0143	0.0157	0.0147
UGS02-001	0.0190	0.0154	0.0160	0.0182	0.0142
UGS02-002	0.0152	0.0172	0.0161	0.0156	0.0148
UGS02-003	0.0135	0.0164	0.0156	0.0145	0.0130
UGS02-004	0.0142	0.0180	0.0166	0.0202	0.0138
平均值	0.0164	0.0166	0.0156	0.0165	0.0140

表2显示，SR算法与BGS算法的计算时间在很大程度上依赖于所处理视频文件的长度；MKFE算法处理时间则随着片段中动作类型的数量不同而有所变化；而新方法与SFMR算法的处理时间均依赖于结果聚类数量与视频序列长度这两个因素，因为两个算法都包含了一个反复调整阶段以优化粘滞函数值与聚类中心节点，即新算法的运行时间长短受聚类簇的形成过程影响。

评价结果度量：图2示出了保真度评价标准的结果。为实现公平的比较，所有参试算法基于同一特征标准，即256HSV色彩直方图。由于实验中所用视频片段均来自一个镜头，则保真度的计算都是全局性的，因此此处不考虑不同镜头的情况，即不再计算局部保真度。

实验说明：提出的方法在所有测试视频中的效果是最好的，特别是在02-001与02-004视频片段上的性能表现更加明显。

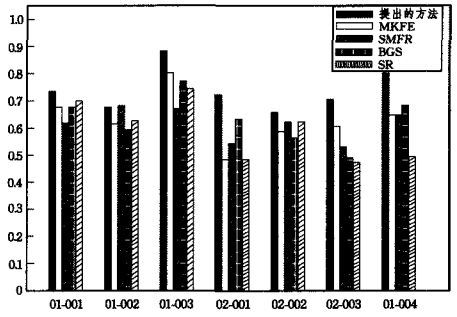

图2 各方法保真度评价

压缩率：表3列出了压缩率结果，通过比较各个算法中提出的正确帧数量及各自的压缩率，可以看出，相较其他方法，新方法提取的关键帧数量较多，其平均压缩率比较高。

表3 各方法的压缩率结果比较

文件名	关键帧数量(CR(%))				
	新方法	MKFE	SMFR	BGS	SR
UGS01-001	27(0.56)	25(0.52)	24(0.50)	25(0.52)	24(0.50)
UGS01-002	18(0.41)	14(0.32)	17(0.38)	14(0.32)	15(0.34)
UGS01-003	24(0.42)	21(0.37)	17(0.30)	20(0.35)	23(0.40)
UGS02-001	16(0.43)	12(0.33)	14(0.38)	15(0.41)	12(0.33)
UGS02-002	17(0.35)	13(0.27)	14(0.29)	12(0.25)	14(0.29)
UGS02-003	17(0.25)	15(0.22)	14(0.20)	13(0.19)	12(0.17)
UGS02-004	21(0.28)	18(0.24)	18(0.24)	19(0.25)	16(0.21)

结束语 本文为大数据的处理提供了一种稀疏图与Grassmann流形测地距的代表选择方法，基于自我表达特性创建了一种数据集的非负稀疏图表示。非负稀疏图方法相比传统的密度图方法更为高效。将流形测地距与min-max算法用于把稀疏图处理成聚类簇，然后从中选择重要的聚类子集，用主成分集中性方法为每个选定的聚类选择一个最终的代表，最后将其运用到视频分析中处理关键帧的提取问题。实验结果表明了所提框架方法的有效性。

参 考 文 献

[1] Dang C, Radha H. Heterogeneity Image Patch Index and Its Application to Consumer Video Summarization[J]. IEEE Transactions on Image Processing, March 2014, 23(6): 2704-2718

[2] Elhamifar, Ehsan, Sapiro G, et al. Finding Exemplars from Pairwise Dissimilarities via Simultaneous Sparse Recovery[J]. Advances in Neural Information Processing Systems, 2012, 1(8): 19-27

[3] Kumar, Mrityunjay, Loui A C. Key frame extraction from consumer videos using sparse representation[C]// Proc. 18th IEEE ICIP. 2011: 2437-2440

[4] Luo Jie-bo, Papin C, Costello K. Towards extracting semantically meaningful key frames from personal video clips: from humans to computers[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2009, 19(2): 289-301

[5] Wang Zhe-shen, Kumar M, Luo Jie-bo, et al. Extracting key frames from consumer videos using bi-layer group sparsity[C]// ACM International Conference on Multimedia. 2011: 1505-1508

（下转第60页）

图 6 中结点的标识如表 2 所列。

表 2　图 6 中结点的标识

标识	标识代表的实际标识
m_0	$C(talk_2, switch_2)$, $Trans_1(talk_1, switch_1, gain_1, lose_1)$, $Car(talk_1, switch_1)$, $ldtrans_2(gain_2, lose_2)$
m_{11}	$Car(talk_1, switch_1)$, $ldtrans_2(gain_2, lose_2)$, $\delta_2(talk_2, switch_2, gain_1, lose_1)$, $\delta_1(talk_1, switch_1, talk_2, switch_2)$
m_{21}	$Car(talk_2, switch_2)$, $ldtrans_2(gain_2, lose_2)$, $ldtrans_1(gain_1, lose_1)$, $\delta_1(talk_1, switch_1, talk_2, switch_2)$
m_{22}	$C(talk_1, switch_1)$, $Car(talk_1, switch_1)$, $\delta_2(talk_2, switch_2, gain_1, lose_1)$, $Trans_2(talk_2, switch_2, gain_2, lose_2)$
m_{31}	$Car(talk_2, switch_2)$, $ldtrans_1(gain_1, lose_1)$, $C(talk_1, switch_1)$, $Trans_2(talk_2, switch_2, gain_2, lose_2)$
m_{32}	$Car(talk_1, switch_1)$, $\delta_2(talk_2, switch_2, gain_1, lose_1)$, $\delta_2'(talk_1, switch_1, gain_2, lose_2)$, $\delta_1'(talk_2, switch_2, talk_1, switch_1)$
m_{41}	$Car(talk_2, switch_2)$, $ldtrans_1(gain_1, lose_1)$, $\delta_2'(talk_1, switch_1, gain_2, lose_2)$, $\delta_1'(talk_2, switch_2, talk_1, switch_1)$
m_{51}	$Car(talk_1, switch_1)$, $ldtrans_1(gain_1, lose_1)$, $ldtrans_2(gain_2, lose_2)$, $\delta_1'(talk_2, switch_2, talk_1, switch_1)$
m_{52}	$Car(talk_2, switch_2)$, $C(talk_2, switch_2)$, $\delta_2'(talk_1, switch_1, gain_2, lose_2)$, $Trans_1(talk_1, switch_1, gain_1, lose_1)$
m_{61}	$Car(talk_2, switch_2)$, $\delta_2'(talk_1, switch_1, gain_2, lose_2)$, $\delta_2(talk_2, switch_2, gain_1, lose_1)$, $\delta_1(talk_1, switch_1, talk_2, switch_2)$

根据图 6 可知，移动车辆电话通信系统的可达树是有限的。

结束语　移动网是在 Petri 网的基础上增加了移动性，并结合了进程代数的优势得到的。移动网改进了 Petri 网结构静态并且无法反映结构变化的弱点，适于模拟和分析移动计算系统。本文给出了移动网的可达树构造算法，为移动网模型提供了基本的可达性分析手段。我们未来将基于移动网的可达性分析方法研究模型的各种动态性质。

参 考 文 献

[1] Asperti A, Busi N. Mobile Petri Nets[J]. Mathematical Structures in Computer Science, 2009; 19(6): 1265-1278

[2] Berry G, Boudol G. The chemical abstract machine[J]. Theoretical Computer Science, 1992, 96(1): 217-248

[3] Du X T. Concurrent computing paradigm: CCS and calculus[J]. Computer Science, 2002, 29(10): 1-4(in Chinese)
杜旭涛. 计算范型：CCS 和 π 演算[J]. 计算机科学, 2002, 29(10): 1-4

[4] Fournet C, Gonthier G. The reflexive cham and the join-calculus [C]// Proceedings of the 23rd Symposium on Principles of Programming Languages. 1996: 372-385

[5] Jensen K. Coloured Petri Nets [M]. EATCS Monographs in Computer Science, SpringerVerlag, 1992: 10-110

[6] Peterson J L. Petri Net Theory and the Modeling of Systems [M]. Englewood Cliffs, NJ: Prentice-Hall, 1981

[7] Milner R, Parrow J, Walker D. A calculus of mobile processes [J]. Information and Computation, 1992, 100(1): 1-77

[8] Montanari U, Rossi F. Contextual nets[J]. Acta Inf., 1995, 32(6): 545-596

[9] Reisig W. An Introduction to Petri Nets [M]. EATCS Monographs in Computer Science, SpringerVerlag, 1985: 1-164

[10] Robin Milner. Communicating and mobile systems: the pi-calculus[M]. Cambridge, the University of Cambridge, 1999: 82-85

[11] Sangiorgi D. Expressing Mobility in Process Algebra[D]. University of Edin-burgh, 1993

[12] Aggarwal S, Pathak H. Modelling of hierarchical location management schemes to locate mobile multi agents using colored Petri Net[M]. Computer Engineering and Applications, 2015: 821-825

[13] Wolfgang Reisig: Understanding Petri Nets[M]. Springer Verlag, 2013: 13-23

[14] 吴哲辉. Petri 网导论[M]. 北京, 2006: 47-193

(上接第 23 页)

[6] Elhamifar, Ehsan, Sapiro G, Vidal R. See all by looking at a few: Sparse modeling for finding representative objects[C]// Computer Vision and Pattern Recognition IEEE Conference on (CVPR). 2012: 1600-1607

[7] Cheng B, Yang J, Yan S, et al. Learning with ℓ_1-graph for image analysis [J]. IEEE Transactions on Image Processing, 2010, 19(4): 858-866

[8] Dang C T, Kumar M, Radha H. Key frame extraction from consumer videos using epitome[C]// Image Processing, International Conference on (ICIP). IEEE, 2012: 93-96

[9] He Ran, Zheng Wei-shi, Hu Bao-gang, et al. Nonegative sparse coding for discriminative semi-supervised learning[C]// Computer Vision and Pattern Recognition (CVPR). CVPR, IEEE, Providence, USA, 2011: 792-801

[10] Wong W K. Discover latent discriminant information for dimensionality reduction: non-negative sparseness preserving embedding [J]. Pattern Recognition, 2012, 45(4): 1511-1523

[11] Lu Gui-fu, Jin Zhong, Zou Jian. Face recognition using discriminant sparsity neighborhood preserving embedding [J]. Knowledge based Systems, 2012, 31(2): 119-127

[12] Shi Jian-bo, Malik J. Normalized Cuts and Image Segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 888-905

[13] Raducanu B, Dornaika F. A supervised non-linear dimensionality reduction approach for manifold learning [J]. Pattern Recognition, 2012, 45(6): 2432-2444

[14] Ilyas M U, Radha H. Identifying Influential Nodes in Online Social Networks Using Principal Component Centrality[C]// 2011 IEEE International Conference on Kyoto. IEEE, Japan, 2011

[15] Almeida, Jurandy, Leite N J, Ricardo da S Torres. Online video summarization on compressed domain [J]. Journal of Visual Communication and Image Representation, 2013, 24(6): 729-738

[16] Almeida, Jurandy, Leite N J, et al. Vision: Video summarization for online applications[J]. Pattern Recognition Letters, 2012, 33(4): 397-409

[17] The Open Video Project[OL]. http://www.open-video.org

[18] Kim Y J, Oh Y T, Yoon S H, et al. Efficient Hausdorff Distance computation for freeform geometric models in close proximity [J]. Computer-Aided Design, 2013, 45(2): 270-276

[19] Schuetze O, Equivel X, Lara A, et al. Some comments on GD and IGD and relations to the Hausdorff distance[C]// Proceedings of the 12th Annual Conference Comp on Genetic and Evolutionary Computation (GECCO's 10:). New York, NY, USA, ACM, 2010: 1971-1974