

DOI: 10.11992/tis.202303049

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20230731.1517.006>

多图融合约束半非负矩阵分解的动作分割方法

李国朋¹, 王连清^{2,3}, 韩鹏¹, 王宇弘⁴, 宋聘¹, 余立^{1,2}

(1. 国防科技大学 试验训练基地, 陕西 西安 710106; 2. 国防科技大学 智能科学学院, 湖南 长沙 410073; 3. 国防科技大学 信息通信学院, 湖北 武汉 430035; 4. 国家工业信息安全发展研究中心, 北京 100040)

摘要: 基于聚类的无监督动作分割方法主要利用序列中相邻帧之间的结构相似性来提高动作分割的准确性。这在实现动作片段内部一致划分的同时给不同动作边界的准确分割带来隐患。为此提出了一种基于多图融合约束矩阵分解的动作分割方法。通过融合序列中的结构相似性和度量相似性信息构造多图融合约束项, 融入到半非负矩阵分解中获得序列的低维表示, 进而获得序列的 k 近邻图并利用图割的方法实现准确分割。在两类动作序列上的实验表明, 所提方法在保持动作内部一致划分的同时能够准确划分动作边界, 明显提升了分割准确性, 时间效率也明显提升。

关键词: 动作分割; 聚类; 半非负矩阵分解; 多图融合约束; 结构相似性; 度量相似性; 低维表示; k 近邻图

中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-4785(2023)06-1223-10

中文引用格式: 李国朋, 王连清, 韩鹏, 等. 多图融合约束半非负矩阵分解的动作分割方法 [J]. 智能系统学报, 2023, 18(6): 1223-1232.

英文引用格式: LI Guopeng, WANG Lianqing, HAN Kun, et al. Action segmentation based on multigraph fusion constraint semi-nonnegative matrix factorization[J]. CAAI transactions on intelligent systems, 2023, 18(6): 1223-1232.

Action segmentation based on multigraph fusion constraint semi-nonnegative matrix factorization

LI Guopeng¹, WANG Lianqing^{2,3}, HAN Kun¹, WANG Yuhong⁴, SONG Dan¹, YU Li^{1,2}

(1. Test Center, National University of Defense Technology, Xi'an 710106, China; 2. College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China; 3. College of Information and Communication, National University of Defense Technology, Wuhan 430035, China; 4. China Industrial Control Systems Cyber Emergency Response Team, Beijing 100040, China)

Abstract: Most clustering-based action segmentation methods mainly exploit the structure similarity information between adjacent frames (points) in the sequence to improve the accuracy of action segmentation. These methods improve the consistency of segmentation inside each action but introduce potential issues for accurately segmenting action boundaries. Hence, this paper presents a novel action segmentation method based on multigraph fusion constraint semi-nonnegative matrix factorization (MGSeNMF). In this method, the structural and measurement similarity information is fused to build a multigraph fusion constraint term, which is fused to semi-NMF to obtain a low-dimensional representation. A k-nearest neighbor graph is also generated for the action sequences, realizing accurate segmentation using the graph cut method. Experimental results on two kinds of real-action datasets show that MGSeNMF can accurately divide the boundary of actions while maintaining consistent segmentation inside each action. Thus, the proposed method improves the accuracy of segmentation and efficiency of running time significantly.

Keywords: action segmentation; clustering; semi-NMF; multigraph fusion constraint; structural similarity; measurement similarity; low-dimensional representation; k-nearest neighbor graph

收稿日期: 2023-03-30. 网络出版日期: 2023-08-01.

基金项目: 国家自然科学基金项目(62101558); 陕西省自然科学基金基础研究计划(2022JM-395); 国防科技大学科研计划项目(ZK21-38).

通信作者: 韩鹏, E-mail: hankun17@nudt.edu.cn.

动作分割是将采集到的动作序列数据分割成若干个不同的动作片段, 其在人体行为理解^[1]、人机交互^[2]、机器人控制^[3]、数据挖掘^[4]等多个领域

具有广泛的应用。与其他数据相比,动作序列数据通常包含多个动作片段。一般认为,序列中相邻帧(点)的变化是连续的、缓慢的,具有较高的相似性^[5]。动作分割时,属于同一个动作片段内的相邻帧应划分到一起;而在不同动作片段边界上的相邻帧应分别划分到对应的动作片段中。因此动作分割的关键包括同一动作片段内所有帧的一致分割和动作片段边界帧的准确分割(即动作边界的准确判断)。目前已有大量工作对该问题进行了深入研究,取得了一系列成果,提出了许多基于监督^[6]或无监督学习^[7]动作分割方法。如基于深度学习模型的方法^[8]通过训练数据集上学到的不同粒度、类型的时序特征,可以实现较好的动作分割效果。然而此类基于监督的方法需要大量标注数据,对计算资源要求也相对较高。相比之下,基于聚类的无监督学习方法将不同动作片段看作一类,动作分割就是将动作序列中所有帧聚类到相应动作片段中。这种方法简单、直观,建模灵活,受到越来越多的重视。

考虑到动作序列中的结构相似性特点,即相邻帧之间是平滑的、缓慢变化的,它们之间具有较高的相似性,因此在基于聚类的分割方法中,特征表示学习过程中融入结构相似性是提升分割准确性的关键。如基于子空间聚类(subspace clustering, SC)^[9], TSC(temporal subspace clustering)^[5]构造了相邻帧之间的约束正则项嵌入到子空间投影过程中; TLRSC(temporal plus low-rank subspace clustering)^[10]在子空间投影中引入结构相似性约束和低秩约束,这些结构相似性的引入可以帮助获得更好的动作序列图表示用于聚类。而基于半非负矩阵分解(semi-nonnegative matrix factorization, Semi-NMF)方法^[11],文献[12]在降维的过程中融入了多近邻帧结构相似性约束,可以获得更准确的动作序列低维特征表示用于后续的聚类。TW-FINCH(temporally-weighted hierarchical clustering)^[13]是一种层次聚类方法,在相似性度量中利用结构最近邻的高相似性来约束相邻帧尽可能的划分到同一动作中。ABD(action boundary detection)^[1]通过寻找动作片段的边界来分割不同动作,同样也利用结构相似性来实现快速分割。显然这些方法的核心思路是利用结构相似性约束相邻帧尽可能一致归属到同类,通过实现动作片段内部聚类的准确性有效提升动作分割的准确性。

然而,结构相似性虽然可以确保动作片段内部的一致归属,但是如果相邻帧处在不同动作片

段的边界上,带来的问题就是会造成边界分割的不准确。这是因为边界处的相邻帧在实际中如果差异确实较大,那么该差异信息是准确判断不同动作片段的依据,而结构相似性约束减小了这种边界处的差异信息,使得边界处不同动作的帧之间相似性关系增强,边界变得更加难以判断,给分割造成困难;并且在结构相似性约束中引入的相邻帧数越多,差异信息减少越多,动作边界越难以判断,导致动作分割的准确性很难进一步提升,这是动作分割的难点^[14]。针对该问题,可以考虑采用其他相似性信息来提升分割效果。如文献[15]利用k近邻相似性约束来提升聚类的准确性,该约束是一种度量相似性,通过使用距离函数计算帧间相似性来寻找前k个高相似性关系,并在特征表示学习中保持该相似性关系。对于动作序列而言,如果边界上的相邻帧差异较大,那么该约束不会影响边界处的差异性,可以减少边界处的错误分割;但是如果不同动作片段中存在一些相似帧(如跑步与跳跃动作中都存在双脚离地的帧),那么该约束可能会将那些本属于不同动作片段的帧归属到同类,造成动作片段内部分割的不一致。可见,对于动作分割而言,不同的相似性约束各有利弊,为了实现准确分割,尽可能保持动作片段内所有帧的相同归属和边界处的准确分割十分有必要。

基于以上分析,不同于当前大多数方法仅利用结构相似性约束来提升动作分割的准确性,为了发挥不同相似性约束的优势,本文提出了融合结构相似性和度量相似性信息的多图融合约束 Semi-NMF(multi-graph constraint Semi-NMF, MG-SeNMF)动作分割方法。该方法首先在特征表示学习过程中,通过融合两种相似性关系图构造正则约束项约束 Semi-NMF 获得序列的低维表示;进而生成动作序列的k近邻图^[16],再使用图割^[17]进行分割。

本文工作可概括如下:

1) 提出一种融合结构相似性和度量相似性信息的多图融合约束 Semi-NMF 动作序列特征表示学习算法,在动作序列的表示中保持更合理、更准确的近邻关系。

2) 提出一种求解多图融合约束 Semi-NMF 优化问题的方法,获得动作序列的低维表示并生成动作序列的k近邻图。

3) 在视频和捕获这两类真实动作序列数据集上进行分割实验。实验结果表明与其他经典方法相比,所提方法适用于不同类型数据,能在保持

动作片段内部一致归属的同时准确分割动作边界, 提升分割准确性, 整体运行效率更高。

1 相关工作

当前基于聚类的动作分割方法主要思路是: 首先在基于 SC、Semi-NMF 等特征表示学习的过程中融入相似性信息, 确保近邻帧之间具有较高相似性; 再利用 k-means、图割等对获得的特征表示进行类划分, 实现分割的目的。其中融入结构相似性约束是此类方法的主要工作。本节主要从基于 Semi-NMF 的特征表示学习和结构相似性约束正则项构造两方面展开论述。

1.1 基于 Semi-NMF 的特征表示学习

设输入动作序列 $\mathbf{X} \in \mathbf{R}^{d \times n}$, 包含 n 个连续的数据帧(点)记作: $\{x_1, x_2, \dots, x_n\}$, 每个数据帧的特征维度为 d 。基于 SC 的各种方法可以获得序列 \mathbf{X} 的关系图表示 $\mathbf{G} \in \mathbf{R}^{n \times n}$, 而基于 Semi-NMF 的方法是在特征空间 $\mathbf{Z} \in \mathbf{R}^{d \times p}$ 上获得一个低维特征表示 $\mathbf{H} \in \mathbf{R}^{p \times n}$, 再利用其他方法获得关系图表示 \mathbf{G} , 其中 p 表示降维后的特征维数。因此, 基于 Semi-NMF 的特征表示学习可描述为一个优化问题:

$$\min_{\mathbf{Z}, \mathbf{H}} J = \|\mathbf{X} - \mathbf{ZH}\|_F^2 \quad \text{s.t. } \mathbf{H} \geq 0 \quad (1)$$

其中 \mathbf{H} 中的元素为非负值。Semi-NMF 能够处理负值数据, 具有建模简单灵活, 可解释性强等特点。

研究表明在 Semi-NMF 过程中施加相似性约束可以提高特征学习的质量^[15,18], 从而实现聚类性能的提升。为了融入这种关系, 通常设近邻关系图 $\mathbf{W} \in \mathbf{R}^{n \times n}$ 描述了原始数据之间的近邻相似性关系, 则约束 Semi-NMF 可描述为

$$\min_{\mathbf{Z}, \mathbf{H}} J = \|\mathbf{X} - \mathbf{ZH}\|_F^2 + \alpha \text{tr}(\mathbf{HLH}^T) \quad \text{s.t. } \mathbf{H} \geq 0 \quad (2)$$

式中 \mathbf{L} 是 \mathbf{W} 的拉普拉斯矩阵。 \mathbf{D} 是对角矩阵, 定义为 $\mathbf{D} = \sum_i \mathbf{W}$, 有 $\mathbf{L} = \mathbf{D} - \mathbf{W}$ 。通过在优化模型中施加此类相似性关系约束, Semi-NMF 在迭代中保持原数据点之间的关系, 从而提高特征学习的性能。特别地, 针对动作序列而言, 设计结构相似性约束正则项融入到特征学习中是序列数据聚类的研究重点。

1.2 结构相似性约束正则项构造

为了在特征表示学习的过程中融入结构相似性, 文献[5]认为在动作关系图 \mathbf{G} 的学习过程中, 当前数据帧 x_i 与它之后的数据帧 x_{i+1} 之间的差异应该尽可能的小。为此可以引入正则约束项 $\|\mathbf{GR}\|_1$, 其中 $\mathbf{R} \in \mathbf{R}^{n \times n}$ 是近邻指示矩阵, 定义如下:

$$\mathbf{R} = \begin{bmatrix} -1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \quad (3)$$

该约束项通过 L_1 范数使得 x_i 与 x_{i+1} 之间的差异尽可能小。基于此约束项, 人们设计了许多基于结构约束聚类的动作分割方法^[19]。文献[12]认为 x_i 与其前后不止一个近邻帧具有相似性, 可以在构造约束项时同时考虑多个近邻帧, 因此可在低维特征表示 \mathbf{H} 学习过程中, 利用 $L_{2,1}$ 范数构造正则约束项 $\|\mathbf{HR}\|_{2,1}$, 其中, 包含 q 个结构近邻的近邻指示矩阵 \mathbf{R} 表示如下:

$$\mathbf{R} = \begin{matrix} \begin{matrix} \uparrow \\ q \\ \downarrow \end{matrix} \end{matrix} \begin{bmatrix} -q_{11} & 0 & 0 & \cdots & 0 \\ 1_{21} & -q_{22} & 0 & \cdots & 0 \\ 1_{31} & 1_{32} & -q_{33} & \cdots & 0 \\ 1_{(q+1)1} & 1_{(q+1)2} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & -q_{(n-q)(n-q)} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1_{n(n-q)} \end{bmatrix} \quad (4)$$

文献[20]认为此类多近邻相似性约束项中, 离序列点越远的点, 其相似性越低, 所占权重应当不同。因此, 在关系图 \mathbf{G} 学习中引入遗忘因子来构造约束项 $\|\mathbf{G} - \mathbf{GR}^T\|_2$, 包含 3 个结构近邻的近邻指示矩阵 \mathbf{R} 可描述为:

$$\mathbf{R} = \begin{bmatrix} \beta & 0 & 0 & \cdots & 0 \\ \alpha\beta & \beta & 0 & \cdots & 0 \\ \alpha^2\beta & \alpha\beta & \beta & \cdots & 0 \\ 0 & \alpha^2\beta & \alpha\beta & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \alpha^2\beta & \cdots & \beta \end{bmatrix} \quad (5)$$

式中: 参数 $\beta = 1 - \alpha$, α 是遗忘因子。相关实验表明, 增加相邻帧的个数, 可以强化动作片段内部的划分一致性, 提升动作分割的准确性。

综上所述, 目前多数研究主要考虑如何设计更合理的结构近邻约束, 忽视了近邻约束给动作边界处的准确分割可能带来的问题。如果要进一步提升动作分割的准确性, 边界处的准确分割也同样重要。因此为了兼顾动作片段内部的一致归属和边界处的分割问题, 本文通过融合不同近邻约束来有效降低目前方法存在的潜在弊端, 实现更准确的动作分割。

2 基于多图融合约束矩阵分解的特征表示

2.1 多图融合约束项构建

2.1.1 两种相似性关系图的构建

设 x_i 是动作序列中的一帧,考虑利用 x_i 前后的各 q 个近邻来构造该点结构近邻约束,则 x_i 的结构近邻集合可描述为

$$C(x_i) = \{x_j | |j-i| \leq q \text{ 且 } j \neq i\} \quad (6)$$

式中: $C(x_i)$ 表示 x_i 的 q 结构近邻集合, i 和 j 分别表示 x_i 和 x_j 在序列中所处的位置,若它们之间所处位置相差不超过 q ,则 x_j 是 x_i 的一个结构近邻。此时序列中所有帧的 q 结构近邻集合构成了序列的结构近邻相似性关系,描述为一个无向无权图 $\mathbf{W}_1 = [V E_1]$ 。其顶点 $V = \{x_1, x_2, \dots, x_n\}$ 是序列中所有帧的集合, E_1 是顶点之间边的集合,定义如下:

$$E_1(x_i, x_j) = \begin{cases} 1, & x_j \in C(x_i) \text{ 或 } x_i \in C(x_j) \\ 0, & \text{其他} \end{cases} \quad (7)$$

同时,对于度量相似性关系,选择经典的 k 近邻进行描述^[15]。采用“binary”距离计算所有帧与 x_i 的距离,距离越小,说明与 x_i 相似性越高,其中距离最小的前 k 个帧就构成了 x_i 的 k 近邻集合,描述为

$$N(x_i) = \{x_j | d_{x_i, x_j} \leq d_{ik}\} \quad (8)$$

式中: d_{x_i, x_j} 表示 x_i 和 x_j 之间的距离, d_{ik} 表示从最短距离开始的第 k 个最短距离。若 $d_{x_i, x_j} \leq d_{ik}$,表明 x_j 是 x_i 的一个 k 近邻帧。同理,序列中所有帧的 k 近邻集合就构成了序列度量相似性关系,描述为无向无权图 $\mathbf{W}_2 = [V E_2]$ 。顶点 V 也是序列中所有帧的集合,而 E_2 是顶点之间边的集合,其定义如下:

$$E_2(x_i, x_j) = \begin{cases} 1, & x_j \in N(x_i) \text{ 或 } x_i \in N(x_j) \\ 0, & \text{其他} \end{cases} \quad (9)$$

2.1.2 两多关系图的融合

如前所述, \mathbf{W}_1 和 \mathbf{W}_2 从不同相似性角度描述了帧间关系,各有利弊,为了充分吸收它们的优点,可以融合这两种相似性信息来实现融合约束,提升特征表示学习的质量。基本思路是采用线性组合的策略^[21],对 \mathbf{W}_1 和 \mathbf{W}_2 进行融合。为此构建多图融合约束项如下:

$$\sum_{i=1}^2 \tau_i \text{tr}(\mathbf{H} \mathbf{L}_i \mathbf{H}^T) \quad (10)$$

其中 \mathbf{L}_i 表示第 i 个关系图的拉普拉斯矩阵,定义为 $\mathbf{L}_i = \mathbf{D}_i - \mathbf{W}_i$, $\mathbf{D}_i = \sum_j \mathbf{W}_{ij}$, τ_i 表示第 i 个关系图在组合中的权重,定义为

$$\tau_i = \frac{\tilde{\tau}_i}{\sum_i \tilde{\tau}_i}, \tilde{\tau}_i = \frac{1}{2 \sqrt{\text{tr}(\mathbf{H} \mathbf{L}_i \mathbf{H}^T)}} \quad (11)$$

显然,在融合约束项中,不同图约束项的权重

取决于各自图差异值计算的结果。计算结果越小,说明相似性越高,该约束项权重越大。考虑到同一动作片段内,相邻帧之间相似性较高,此时 x_{i+1} 既是 x_i 的结构近邻又是度量近邻,那么融合约束项可以有效保持它们之间的近邻关系;而在边界处,若 x_{i+1} 不是 x_i 的度量近邻,那么图 \mathbf{W}_1 计算所得相似性结果要小于 \mathbf{W}_2 计算结果, \mathbf{W}_1 在融合约束项中权重小于 \mathbf{W}_2 权重,因此结构相似性约束力度将会减小,从而有利于边界处差异性特征的保留。

2.2 多图融合约束 Semi-NMF 低维特征学习

对于输入序列,将上述多图融合约束项融入到 Semi-NMF 中,构建低维特征学习优化模型如下所示:

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{H}} J &= \|\mathbf{X} - \mathbf{Z}\mathbf{H}\|_F^2 + \alpha \sum_{i=1}^2 \tau_i \text{tr}(\mathbf{H} \mathbf{L}_i \mathbf{H}^T) \\ \text{s.t. } \mathbf{H} &\geq 0, \tau_i = \frac{\tilde{\tau}_i}{\sum_i \tilde{\tau}_i}, \tilde{\tau}_i = \frac{1}{2 \sqrt{\text{tr}(\mathbf{H} \mathbf{L}_i \mathbf{H}^T)}} \end{aligned} \quad (12)$$

式中: α 是多图融合约束项的权重参数, $\|\cdot\|_F^2$ 表示 L_2 范数, τ 为不同关系图在多图融合约束项中所占权重。求解该优化模型是一个非凸优化问题,根据文献[11],可采用交替迭代的思路获得局部最优解。具体方法如下:

1) 更新 \mathbf{Z}

固定其他变量,由于 \mathbf{Z} 可以取负值,此时根据矩阵运算规则,可得 \mathbf{Z} 的闭式解为

$$\mathbf{Z} = \mathbf{X} \mathbf{H}^T (\mathbf{H} \mathbf{H}^T)^{-1} \quad (13)$$

2) 更新 \mathbf{H}

固定其他变量, \mathbf{H} 的求解转化为关于 \mathbf{H} 的如下问题:

$$\begin{aligned} \min L(\mathbf{H}) &= \text{tr}(\mathbf{X} \mathbf{X}^T - 2 \mathbf{X}^T \mathbf{Z} \mathbf{H} + \mathbf{H}^T \mathbf{Z}^T \mathbf{Z} \mathbf{H}) + \\ &\alpha \sum_{i=1}^2 \tau_i \text{tr}(\mathbf{H} \mathbf{D}_i \mathbf{H}^T - \mathbf{H} \mathbf{W}_i \mathbf{H}^T) \end{aligned} \quad (14)$$

考虑到 \mathbf{X} 和 \mathbf{Z} 均可取负值,不失一般地,令 \mathbf{M}^+ 、 \mathbf{M}^- 分别表示矩阵 \mathbf{M} 非负部分和负值部分,即: $\mathbf{M}^+ = (|\mathbf{M}| + \mathbf{M})/2$, $\mathbf{M}^- = (|\mathbf{M}| - \mathbf{M})/2$,于是问题可改写为

$$\begin{aligned} \min L(\mathbf{H}) &= \text{tr}(\mathbf{X} \mathbf{X}^T - 2(\mathbf{X}^T \mathbf{Z})^+ \mathbf{H} + 2(\mathbf{X}^T \mathbf{Z})^- \mathbf{H}) + \\ &\text{tr}(\mathbf{H}^T (\mathbf{Z}^T \mathbf{Z})^+ \mathbf{H} - \mathbf{H}^T (\mathbf{Z}^T \mathbf{Z})^- \mathbf{H}) + \\ &\alpha \sum_{i=1}^2 \tau_i \text{tr}(\mathbf{H} \mathbf{D}_i \mathbf{H}^T - \mathbf{H} \mathbf{W}_i \mathbf{H}^T) \end{aligned} \quad (15)$$

根据 Semi-NMF 的迭代规则求解理论^[11], \mathbf{H} 的更新规则推导可以通过构造辅助函数并求解该函数来完成。 $L(\mathbf{H})$ 的辅助函数 $G(\mathbf{H}, \mathbf{H}')$ 需满足 $G(\mathbf{H}, \mathbf{H}') \geq L(\mathbf{H})$ 且 $G(\mathbf{H}, \mathbf{H}) = L(\mathbf{H})$ 。可以定义 $L(\mathbf{H})$

的辅助函数为:

$$G(\mathbf{H}, \mathbf{H}') = \sum_{ij} X_{ij} X_{ij} - 2 \sum_{ij} (\mathbf{Z}^T \mathbf{X})_{ij}^+ \mathbf{H}'_{ij} \left(1 + \log \frac{\mathbf{H}_{ij}}{\mathbf{H}'_{ij}} \right) + \sum_{ij} (\mathbf{Z}^T \mathbf{X})_{ij}^- \left(\frac{\mathbf{H}_{ij}^2 + \mathbf{H}_{ij}'^2}{\mathbf{H}'_{ij}} \right) + \sum_{ij} ((\mathbf{Z}^T \mathbf{Z})^+ \mathbf{H}')_{ij} \frac{\mathbf{H}_{ij}^2}{\mathbf{H}'_{ij}} - \sum_{ijl} ((\mathbf{Z}^T \mathbf{Z})_{jl}^- \mathbf{H}'_{ij} \mathbf{H}'_{jl}) \left(1 + \log \frac{\mathbf{H}_{ij} \mathbf{H}_{il}}{\mathbf{H}'_{ij} \mathbf{H}'_{il}} \right) + \alpha \sum_{k=1}^2 \tau_k \left(\sum_{ij} (\mathbf{H}'^T \mathbf{D}_k)_{ij} \frac{\mathbf{H}_{ij}^2}{\mathbf{H}'_{ij}} - \sum_{ijl} (\mathbf{W}_k)_{jl} \mathbf{H}'_{ij} \mathbf{H}'_{jl} \left(1 + \log \frac{\mathbf{H}_{ij} \mathbf{H}_{il}}{\mathbf{H}'_{ij} \mathbf{H}'_{il}} \right) \right)$$

首先需要证明 $G(\mathbf{H}, \mathbf{H}')$ 是 $L(\mathbf{H})$ 的一个辅助函数。

证明 显然 $G(\mathbf{H}, \mathbf{H}) = L(\mathbf{H})$ 满足, 重点证明 $G(\mathbf{H}, \mathbf{H}') \geq L(\mathbf{H})$ 。

当非负矩阵 $\mathbf{S} \in \mathbf{R}^{n \times n}$ 、 $\mathbf{B} \in \mathbf{R}^{p \times p}$ 是对角矩阵时, 如下不等式成立:

$$\text{tr}(\mathbf{H}^T \mathbf{S} \mathbf{H} \mathbf{B}) \leq \sum_{i=1}^n \sum_{j=1}^p (\mathbf{S} \mathbf{H}^T \mathbf{B})_{ij} \frac{\mathbf{H}_{ij}^2}{\mathbf{H}'_{ij}} \quad (16)$$

利用该不等式, 可得:

$$\text{tr}(\mathbf{H}^T (\mathbf{Z}^T \mathbf{Z})^+ \mathbf{H}) \leq \sum_{ij} ((\mathbf{Z}^T \mathbf{Z})^+ \mathbf{H}')_{ij} \frac{\mathbf{H}_{ij}^2}{\mathbf{H}'_{ij}} \quad (17)$$

$$\sum_{k=1}^2 \tau_k \text{tr}(\mathbf{H} \mathbf{D}_k \mathbf{H}^T) \leq \sum_{k=1}^2 \tau_k \sum_{ij} (\mathbf{H}'^T \mathbf{D}_k)_{ij} \frac{\mathbf{H}_{ij}^2}{\mathbf{H}'_{ij}} \quad (18)$$

利用不等式 $z > 1 + \log z$ ($z > 0$), 可得:

$$\text{tr}((\mathbf{X}^T \mathbf{Z})^+ \mathbf{H}) \geq \sum_{ij} (\mathbf{Z}^T \mathbf{X})_{ij}^+ \mathbf{H}'_{ij} \left(1 + \log \frac{\mathbf{H}_{ij}}{\mathbf{H}'_{ij}} \right) \quad (19)$$

$$\begin{aligned} \text{tr}(\mathbf{H}^T (\mathbf{Z}^T \mathbf{Z})^- \mathbf{H}) &\geq \\ \sum_{ijl} ((\mathbf{Z}^T \mathbf{Z})_{jl}^- \mathbf{H}'_{ij} \mathbf{H}'_{jl}) &\left(1 + \log \frac{\mathbf{H}_{ij} \mathbf{H}_{il}}{\mathbf{H}'_{ij} \mathbf{H}'_{il}} \right) \end{aligned} \quad (20)$$

$$\begin{aligned} \sum_{k=1}^2 \tau_k \text{tr}(\mathbf{H} \mathbf{W}_k \mathbf{H}^T) &\geq \\ \sum_{k=1}^2 \tau_k \sum_{ijl} (\mathbf{W}_k)_{jl} \mathbf{H}'_{ij} \mathbf{H}'_{jl} &\left(1 + \log \frac{\mathbf{H}_{ij} \mathbf{H}_{il}}{\mathbf{H}'_{ij} \mathbf{H}'_{il}} \right) \end{aligned} \quad (21)$$

利用该不等式 $a^2 + b^2 \geq 2ab$ ($a, b > 0$), 可得:

$$2 \text{tr}((\mathbf{X}^T \mathbf{Z})^- \mathbf{H}) \leq \sum_{ij} (\mathbf{Z}^T \mathbf{X})_{ij}^- \left(\frac{\mathbf{H}_{ij}^2 + \mathbf{H}_{ij}'^2}{\mathbf{H}'_{ij}} \right) \quad (22)$$

综上所述可得: $G(\mathbf{H}, \mathbf{H}') \geq L(\mathbf{H})$, 因此它是 $L(\mathbf{H})$ 的辅助函数。

证毕。

其次通过求解 $G(\mathbf{H}, \mathbf{H}')$ 来求解优化问题 (14), 考虑到 $G(\mathbf{H}, \mathbf{H}')$ 是一个凸优化, 令:

$$\frac{\partial G(\mathbf{H}, \mathbf{H}')}{\partial \mathbf{H}_{ij}} = 0 \quad (23)$$

可得 \mathbf{H} 迭代规则:

$$\mathbf{H} \leftarrow \mathbf{H} \sqrt{\frac{(\mathbf{Z}^T \mathbf{X})^+ + (\mathbf{Z}^T \mathbf{Z})^- \mathbf{H} + \alpha \sum_{i=1}^2 \tau_i \mathbf{H} \mathbf{W}_i}{(\mathbf{Z}^T \mathbf{X})^- + (\mathbf{Z}^T \mathbf{Z})^+ \mathbf{H} + \alpha \sum_{i=1}^2 \tau_i \mathbf{H} \mathbf{D}_i}} \quad (24)$$

3) 更新 τ

为了确定不同关系图的权重, 需要根据每次迭代过程中 \mathbf{H} 的重新确定 \mathbf{W}_1 和 \mathbf{W}_2 的权重系数。因此迭代中 τ 需要按照式 (11) 重新计算权重。

通过以上的迭代更新, 最终可以获得动作序列的低维特征表示 \mathbf{H} , 利用该特征表示可以获取序列的 k 近邻图表示。

3 基于 k 近邻图的动作分割

考虑到获得的矩阵 \mathbf{H} 中蕴含丰富的各种帧间相似性关系, 可以清晰地通过这种相似性关系描述整个动作序列的关系图, 因此矩阵 \mathbf{H} 生成 k 近邻图 $\mathbf{G} \in \mathbf{R}^{n \times n}$ 是一个无向无权图 $\mathbf{G} = [\mathbf{V} \ \mathbf{E}]$, 顶点 \mathbf{V} 是序列中所有帧的集合, \mathbf{E} 是顶点之间边的集合, 定义如下:

$$E(h_i, h_j) = \begin{cases} 1, & h_j \in N(h_i) \text{ 或 } h_i \in N(h_j) \\ 0, & \text{其他} \end{cases} \quad (25)$$

其中 $N(h_i)$ 表示利用特征表示矩阵 \mathbf{H} 计算获取的第 i 帧的 k 近邻集合, k 近邻计算同样按照文献 [14] 方法进行。

最后利用图割方法对 \mathbf{G} 进行分割, 获得动作分割结果。Ncut 是常用的图分割方法, 其利用 \mathbf{D} 的拉普拉斯特征值第 2 小所对应的特征向量来分割, 详细介绍可参考文献 [17]。

综上, 本文提出的 MGSeNMF 动作分割方法可描述如下:

算法 基于 MGSeNMF 动作分割

输入 \mathbf{X}, q, k, α

1) 初始化:

$(\mathbf{Z}, \mathbf{H}) \leftarrow \text{Semi-NMF}(\mathbf{X}), \mathbf{W}_1 \leftarrow \text{式 (6~7)},$

$\mathbf{W}_2 \leftarrow \text{式 (8~9)}, \tau_1 = \tau_2 = 0.5$

2) while 没有收敛 do

$\mathbf{Z} \leftarrow \text{式 (13)}, \mathbf{H} \leftarrow \text{式 (24)}, \tau \leftarrow \text{式 (11)}$

3) end while

4) 生成 k 近邻图: $\mathbf{G} \leftarrow \mathbf{H}$

5) 聚类: Ncut(\mathbf{G})

输出 分割结果

其中初始矩阵 \mathbf{Z} 、 \mathbf{H} 采用经典 Semi-NMF 进行赋值, τ 初始值都是 0.5 表示两种图关系的权重相同。

4 实验与讨论

4.1 数据集及比较方法

实验在两类不同的真实动作序列共 6 个数据集上进行。一类是视频序列, 包括 Weizmann (Weiz)、Keck 和 UT-Interaction (UT-I)^[22]。其中 Weiz 和 UT-I 是人体动作序列, Keck 是手势动作序列。与文献^[22]一样, 提取 HoG 特征作为这些序列的输入数据。另一类来自 CMU Mocap database 86 号运动捕获序列数据^[23]。这些运动序列主要由各个关节附着传感器的人在一定时间内做出不同的动作而生成, 共 14 组。选取其中的第 1、2、3 组, 记作: mo-1、mo-2、mo-3。这些数据集信息如表 1 所示。

表 1 实验数据集信息
Table 1 Characteristics of datasets

数据集	序列帧数	特征维度	动作数	重复动作
Weiz	701	324	10	无
Keck	1245	324	10	无
UT-I	650	1024	6	无
mo-1	954	42	4	有
mo-2	2192	42	8	有
mo-3	1760	42	7	有

为了比较不同方法动作分割的效果, 与 TSC、TLRSC、SC-NMF、TW-FINCH 等基于聚类的动作分割方法进行比较。其中: TSC 和 TLRSC 是基于 SC 的动作分割方法; SC-NMF 是基于 Semi-NMF 的方法, 它在低维表示过程中引入了 q 个结构近邻构造的约束项, 再生成一个 k 近邻图用于动作分割; TW-FINCH 是一个基于层次聚类的动作分割方法。

实验的参数分别根据文献的推荐进行设置。其中 MGSeNMF 与 SC-NMF 一样, 设置 Weiz、Keck、UT-I 的特征维数为 60, mo-1、mo-2、mo-3 的特征维数为 15; α 取值范围 [0.001, 0.01, 0.1, 1, 10, 100, 1000]。MGSeNMF 设置 q 取值范围 [1, 2, 3, 4, 5, 6, 8, 10], k 与 q 相同。SC-NMF 中 q 的取值与文献保持一致, 为 MGSeNMF 中 q 的 2 倍。所有方法获取最优结果作为输出。评价指标采用标准的 ACC 和 NMI 进行衡量^[24-25]。

4.2 动作分割结果比较

表 2 和表 3 所示是不同方法动作分割的 ACC 和 NMI 性能比较。可以看到, MGSeNMF 在大多数数据集上都获得了更好的动作分割效果, 特别是在 3 个视频序列数据集上性能提升效果明显; 在 mo-1 和 mo-3 上, MGSeNMF 也可以获得最好的 ACC 和 NMI 值, 在 mo-2 上稍差于 SC-NMF。

为进一步分析动作分割效果, 图 1~6 所示为不

同方法的最终动作分割结果(不同颜色表示不同的动作片段)比较, 其中 GND 表示数据集的标准结果。

表 2 不同方法的 ACC 性能比较
Table 2 ACC comparison with different methods

数据集	TSC	TLRSC	SC-NMF	TW-FINCH	MGSeNMF
Weiz	0.7874	0.7789	0.8759	0.7660	0.9244
Keck	0.9044	0.8000	0.9173	0.7406	0.9566
UT-I	0.8738	0.8277	0.9369	0.5800	0.9534
mo-1	0.6709	0.8899	0.9644	0.6583	0.9644
mo-2	0.8244	0.8166	0.9630	0.7865	0.9576
mo-3	0.7364	0.7432	0.9489	0.8051	0.9517

表 3 不同方法的 NMI 性能比较
Table 3 NMI comparison with different methods

数据集	TSC	TLRSC	SC-NMF	TW-FINCH	MGSeNMF
Weiz	0.8303	0.8116	0.8769	0.8279	0.9468
Keck	0.8799	0.8689	0.8939	0.8231	0.9389
UT-I	0.8112	0.7255	0.8937	0.5057	0.9336
mo-1	0.6515	0.6822	0.8693	0.6021	0.8733
mo-2	0.8007	0.7107	0.9119	0.7498	0.9021
mo-3	0.7561	0.6757	0.8698	0.7465	0.8786

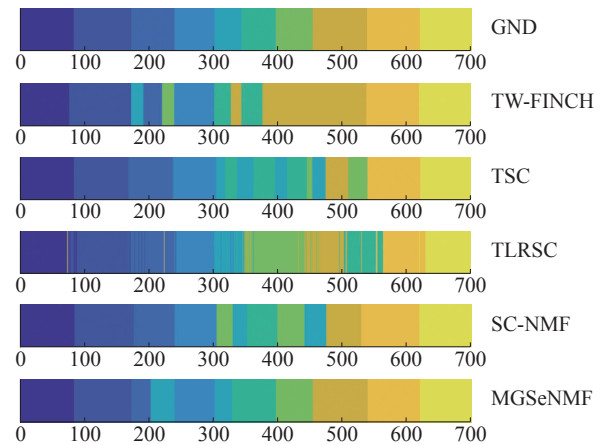


图 1 Weiz 数据集上的动作分割结果

Fig. 1 The action segmentation yielded on Weiz

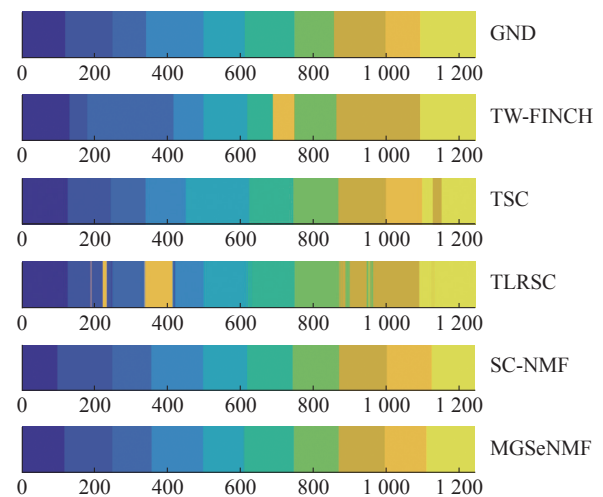


图 2 Keck 数据集上的动作分割结果

Fig. 2 The action segmentation yielded on Keck

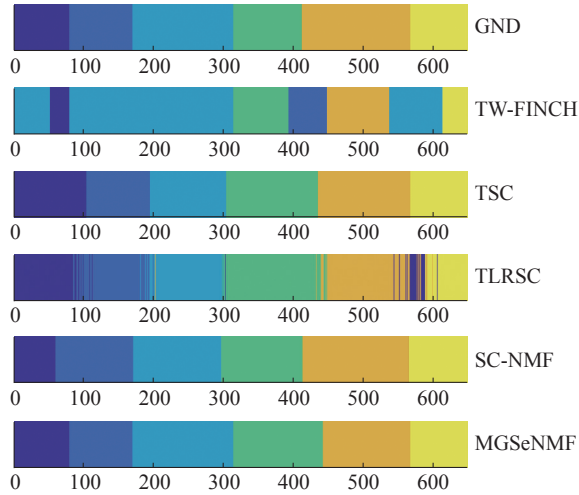


图 3 UT-I 数据集上的动作分割结果

Fig. 3 The action segmentation yielded on UT-I

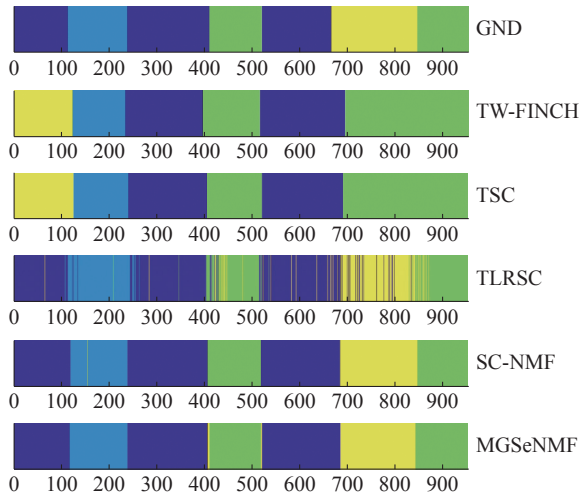


图 4 mo-1 数据集上的动作分割结果

Fig. 4 The action segmentation yielded on mo-1

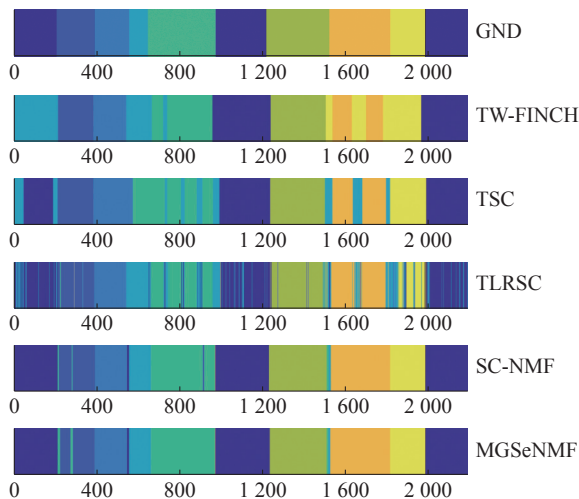


图 5 mo-2 数据集上的动作分割结果

Fig. 5 The action segmentation yielded on mo-2

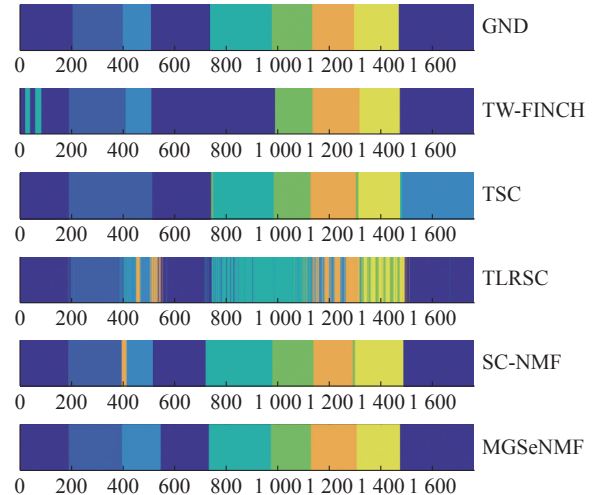


图 6 mo-3 数据集上的动作分割结果

Fig. 6 The action segmentation yielded on mo-3

可以看到,与其他方法相比,MGSeNMF 与 SC-NMF 基本都可以保持动作序列内部的一致归属。它们的主要区别在于动作边界处的分割结果。由于 MGSeNMF 引入 k 近邻图信息,在保持有效近邻关系的同时,弱化了结构近邻图的强制性约束,从而可以在动作边界处实现更客观准确的分割。如在 Weiz 数据集上,虽然 MGSeNMF 将动作 3 的结束部分划分到动作 5,但是主要动作之间的边界都是清晰的,与标准分割一致,没有出现如 SC-NMF 那样将动作 7 的结束和动作 8 的开始分割为一个独立新动作之类的错误。Keck 数据集上,动作 1 和动作 2、动作 9 和动作 10 之间的分割更接近标准。可见利用多图融合信息,MGSeNMF 在确保动作片段内一致归属的同时,边界处的准确分割得到了进一步增强,从而使得动作分割的准确性得到了进一步提升。

4.3 参数分析

MGSeNMF 中参数 k 和 q 分别描述两种相似图中近邻的个数,实验取值在 $[1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 20]$ 范围内,如图 7~12 所示为不同取值下的 NMI 性能。可以看到,同时增大 k 和 q ,NMI 性能在 Weiz、Keck、UT-I、mo-3 上提升明显,说明引入较多的结构近邻节点,可以更有效约束动作序列结构近邻之间表示尽可能的一致,从而确保动作片段内部归属到相同类。同时需要注意到,过大的 q 仍意味着会将不同动作边界处的数据帧建立高相似性关系,从而导致错误分割。如当 $q=20$ 时,NMI 性能在 Weiz、Keck、UT-I、mo-1、mo-3 上下降明显,在其他数据集上也有下降,因此 q 的选择不能过大。同理对于 k 的取值,和文献 [14] 所得结论一样,过大的 k 也将导致性能的下降,因此对 k 的讨论本文不再赘述。

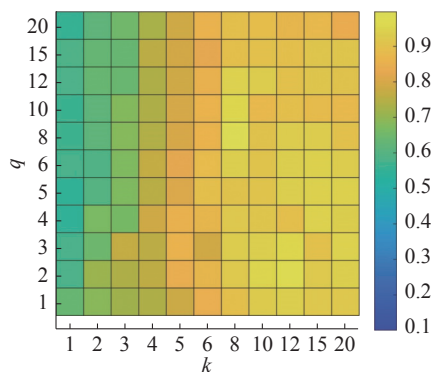


图7 Weiz数据集上不同 q 和 k 下的NMI值
Fig. 7 The NMI with different q and k on Weiz

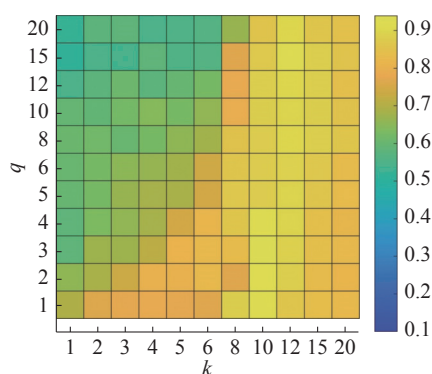


图8 Keck数据集上不同 q 和 k 下的NMI值
Fig. 8 The NMI with different q and k on Keck

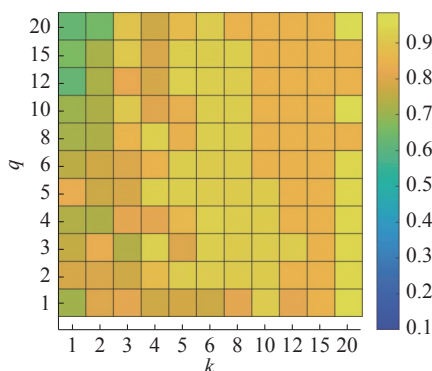


图9 UT-I数据集上不同 q 和 k 下的NMI值
Fig. 9 The NMI with different q and k on UT-I

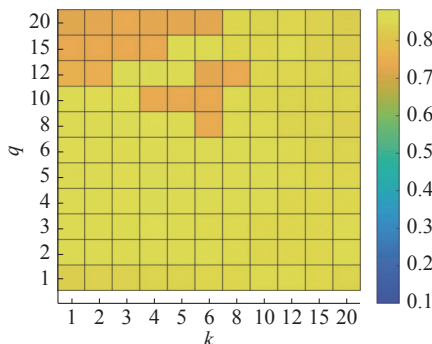


图10 mo-1数据集上不同 q 和 k 下的NMI值
Fig. 10 The NMI with different q and k on mo-1

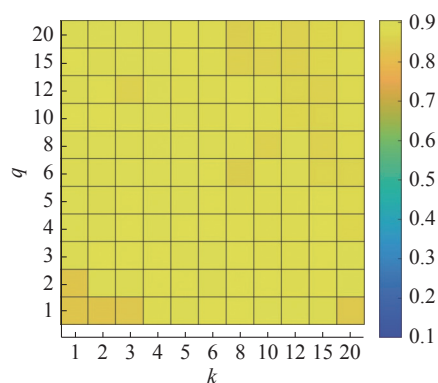


图11 mo-2数据集上不同 q 和 k 下的NMI值
Fig. 11 The NMI with different q and k on mo-2

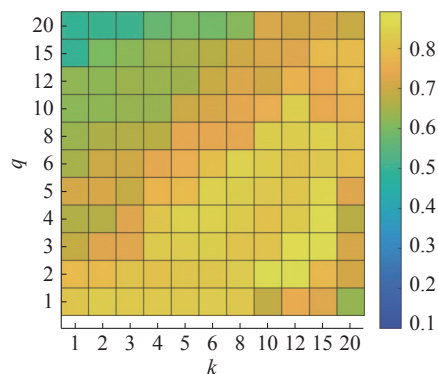


图12 mo-3数据集上不同 q 和 k 下的NMI值
Fig. 12 The NMI with different q and k on mo-3

MGS_{se}NMF 中参数 α 描述了多图融合约束项在优化目标中的权重。固定其他参数, 设置 α 取值 $[0.001, 0.01, 0.1, 1, 10, 100, 1000]$, 以 Weiz 和 mo-1 为例, 图 13~14 所示为不同 α 取值下的 ACC 和 NMI 值。可以看到, 在 Weiz 数据集上, 当 $\alpha \leq 1$ 时, ACC 和 NMI 提升明显, 之后略有下降; 与之相反, 在 mo-1 数据集上, 当 $\alpha \geq 0.1$ 时, ACC 和 NMI 下降明显。这是因为对于 Weiz 之类的高维数据而言, 需要较大的 α 值来提升多图融合约束项在整个优化模型中的权重; 反之对于 mo-1 等数据而言, 由于原始数据特征维数相对较小, 因此需要较小的 α 值来调节多图融合约束项在优化模型中的权重。

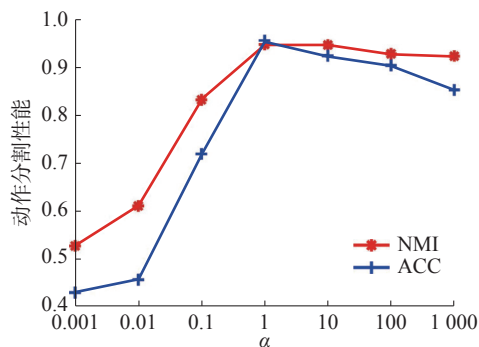
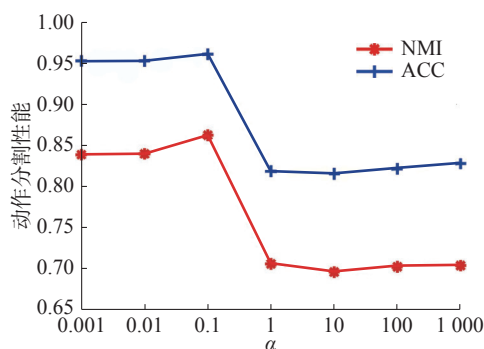


图13 Weiz数据集上ACC和NMI值随 α 变化情况
Fig. 13 The ACC and NMI with different set of α on Weiz

图14 mo-1数据集上ACC和NMI值随 α 变化情况Fig. 14 The ACC and NMI with different set of α on mo-1

4.4 运行时间分析

表4为不同方法所需时间的统计。可以看到, TW-FINCH所需时间最少, SC-NMF方法最耗时, TLRSC次之。结合前述分割性能分析可以看到, 在整体上取得更好分割准确性的情况下, MGSeNMF运行所需时间大大缩小, 运行效率明显提升。这说明虽然增加了一个图约束, 但是MGSeNMF的多图融合约束项设计采用 L_2 范数, 计算比SC-NMF中的 $L_{2,1}$ 更高效, 因此MGSeNMF整体计算效率更高。特别是随着数据量的增大, MGSeNMF运行效率更具优势。

表4 不同方法的运行时间比较

Table 4 The comparison of the running time with different methods

数据集	TSC	TLRSC	SC-NMF	TW-FINCH	MGSeNMF
Weiz	0.8683	1.3697	5.6972	0.4058	2.4358
Keck	1.6168	4.8103	24.3205	0.4389	5.1995
UT-I	0.8986	1.0194	5.2796	0.7519	2.7492
mo-1	1.1796	1.7360	10.5217	0.3990	2.9585
mo-2	2.9823	20.9330	113.3782	0.4305	6.9445
mo-3	2.2592	12.1045	60.1259	0.4135	4.9526
合计	9.8048	41.9729	219.3231	2.8396	25.2401

5 结束语

针对当前基于聚类的动作分割方法常采用结构相似性约束易导致边界上的错误而造成动作分割不准确的问题, 提出了一种融合结构相似性和度量相似性的基于多图融合约束 Semi-NMF 的动作分割方法 MGSeNMF。该方法主要在特征表示学习中, 利用不同权重的相似性关系图融合约束实现序列的降维, 从而确保了后续获得更好的序列关系图表示和更准确的动作分割。实验表明该方法可以有效提升动作分割的准确性, 特别是在视频序列中, 平均精度较第二名方法 SC-NMF 提升了近 4%; 并且在准确性得到提高的同时, 运行效率明显提升, 与 SC-NMF 相比, 总的运行时间

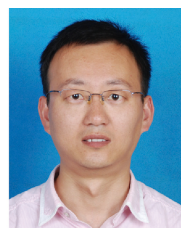
缩短了近一个数量级。下一步, 针对自动设置最优近邻个数, 更准确确定动作边界以及运用多模态数据实现更优动作分割等问题持续开展研究。

参考文献:

- [1] DU Zexing, WANG Xue, ZHOU Guoqing, et al. Fast and unsupervised action boundary detection for action segmentation[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 3313–3322.
- [2] CHAKRABORTY B K, SARMA D, BHUYAN M K, et al. Review of constraints on vision-based gesture recognition for human-computer interaction[J]. *IET computer vision*, 2018, 12(1): 3–15.
- [3] 邵振洲, 赵红发, 渠瀛, 等. 一种面向多模态手术轨迹的快速无监督分割方法[J]. *小型微型计算机系统*, 2018, 39(10): 2296–2302.
- [4] SHAO Zhenzhou, ZHAO Hongfa, QU Ying, et al. Fast unsupervised approach for multi-modality surgical trajectory segmentation[J]. *Journal of Chinese computer systems*, 2018, 39(10): 2296–2302.
- [4] CHEN Wanjun, ZHANG Erhu, ZHANG Zhuomin. A Laplacian structured representation model in subspace clustering for enhanced motion segmentation[J]. *Neuro-computing*, 2016, 208: 174–182.
- [5] LI Sheng, LI Kang, FU Yun. Temporal subspace clustering for human motion segmentation[C]//2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2016: 4453–4461.
- [6] GAO Shanghua, HAN Qi, LI Zhongyu, et al. Global2Local: efficient structure search for video action segmentation[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 16800–16809.
- [7] ZHOU Tao, FU Huazhu, GONG Chen, et al. Consistency and diversity induced human motion segmentation[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2023, 45(1): 197–210.
- [8] LEI Peng, TODOROVIC S. Temporal deformable residual networks for action segmentation in videos[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 6742–6751.
- [9] ELHAMIFAR E, VIDAL R. Sparse subspace clustering: algorithm, theory, and applications[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2013, 35(11): 2765–2781.
- [10] ZHENG Jianwei, YANG Ping, SHEN Guojian, et al. Enhanced low-rank constraint for temporal subspace clustering and its acceleration scheme[J]. *Pattern recognition*,

- 2021, 111: 107678.
- [11] DING C H Q, LI Tao, JORDAN M I. Convex and semi-nonnegative matrix factorizations[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2010, 32(1): 45–55.
- [12] GAO Hongbo, LYU Chen, ZHANG Tong, et al. A structure constraint matrix factorization framework for human behavior segmentation[J]. *IEEE transactions on cybernetics*, 2022, 52(12): 12978–12988.
- [13] SARFRAZ M S, MURRAY N, SHARMA V, et al. Temporally-weighted hierarchical clustering for unsupervised action segmentation[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 11220–11229.
- [14] WANG Xiumei, GUO Dingning, CHENG Peitao. Support structure representation learning for sequential data clustering[J]. *Pattern recognition*, 2022, 122: 108326.
- [15] CAI Deng, HE Xiaofei, HAN Jiawei, et al. Graph regularized nonnegative matrix factorization for data representation[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2011, 33(8): 1548–1560.
- [16] CAI Yongda, HUANG J Z, YIN Jianfei. A new method to build the adaptive k-nearest neighbors similarity graph matrix for spectral clustering[J]. *Neurocomputing*, 2022, 493: 191–203.
- [17] SHI Jianbo, MALIK J. Normalized cuts and image segmentation[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2000, 22(8): 888–905.
- [18] HE Yangcheng, LU Hongtao, HUANG Lei, et al. Pairwise constrained concept factorization for data representation[J]. *Neural networks*, 2014, 52: 1–17.
- [19] XIA Guiyu, CHEN Beijia, SUN Huaijiang, et al. Nonconvex low-rank kernel sparse subspace learning for key-frame extraction and motion segmentation[J]. *IEEE transactions on neural networks and learning systems*, 2021, 32(4): 1612–1626.
- [20] LIU Haijun, CHENG Jian, WANG Feng. Sequential subspace clustering via temporal smoothness for sequential data segmentation[J]. *IEEE transactions on image processing*, 2018, 27(2): 866–878.
- [21] ZHANG Kexin, ZHAO Xuezhan, PENG Siyuan. Multiple graph regularized semi-supervised nonnegative matrix factorization with adaptive weights for clustering[J]. *Engineering applications of artificial intelligence*, 2021, 106: 104499.
- [22] WANG Lichen, DING Zhengming, FU Yun. Low-rank transfer human motion segmentation[J]. *IEEE transactions on image processing*, 2019, 28(2): 1023–1034.
- [23] JIANG Zhuolin, LIN Zhe, DAVIS L. Recognizing human actions by learning and matching shape-motion prototype trees[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2012, 34(3): 533–547.
- [24] ZHU Xiaofeng, ZHANG Shichao, HE Wei, et al. One-step multi-view spectral clustering[J]. *IEEE transactions on knowledge and data engineering*, 2019, 31(10): 2022–2034.
- [25] BAI Yue, WANG Lichen, LIU Yunyu, et al. Human motion segmentation via velocity-sensitive dual-side auto-encoder[J]. *IEEE transactions on image processing*, 2023, 32: 524–536.

作者简介:



李国朋, 助理研究员, 博士, 主要研究方向为模式识别、多视图学习及其在无人驾驶智能感知与决策中的应用。



王连清, 博士研究生, 主要研究方向为视频处理与理解、指挥与决策。



韩鹏, 副教授, 主要研究方向为无人系统安全、多智能体博弈。主持、参与国家和省部级科研项目 10 余项, 获省部级科技进步三等奖 1 项, 省部级教学成果三等奖 1 项, 发表学术论文 20 余篇。