

基于 LLE 的视频摘要提取方法研究

冯 燕^{1,2}, 王洪元^{1,2}, 程起才^{1,2}

(1 常州大学 信息科学与工程学院, 江苏 常州 213164; 2 常州市过程感知与互联技术重点实验室, 江苏 常州 213164)

摘 要: 为了自动获取主要视频信息且冗余信息较少的视频摘要, 本文提出了 LLE-自适应 FCM 和 LLE-自适应阈值 FCM 算法. 这两种方法首先利用流形学习算法局部线性嵌入(LLE)提取视频帧的特征向量, 然后将得到的特征向量输入到自适应 FCM 和自适应阈值 FCM 中, 得出分类效果和聚类中心. 自适应 FCM 通过聚类有效性函数来确定分类类别数, 而自适应阈值 FCM 是通过阈值的自动变化来确定分类类别数. 最后把离聚类中心最近的视频帧作为视频摘要. 实验的结果表明, 在不需要人工干预的情况下, 所提取的视频摘要既反映了视频的主要内容, 而且冗余信息少.

关键词: 视频摘要; 流形学习算法; 局部线性嵌入; 自适应 FCM; 自适应阈值 FCM

中图分类号: TP18

文献标识码: A

文章编号: 1000-7180(2012)03-0060-04

Video Abstract Based on Locally Linear Embedding

FENG Yan, WANG Hong-yuan, CHENG Qi-cai

(1 School of Information Science and Engineering, Changzhou University, Changzhou 213164, China;

2 Changzhou Key Laboratory for Process Perception and Interconnected Technology, Changzhou 213164, China)

Abstract: In order to get the main information automatically, while the redundant information is less, the LLE-Adaptive FCM and LLE-Adaptive threshold FCM are presented in this paper. At the beginning of two algorithms, getting the feature vectors of video frames via locally linear embedding (LLE), and then inputting the feature vectors into LLE-Adaptive FCM and LLE-Adaptive threshold FCM, the results of clustering and Clustering centers will be got. Adaptive FCM through clustering validity function to determine the classification numbers, and Adaptive threshold FCM through the automatic changes of threshold to determine the number of classification. Finally, computing the Video abstracts which have the shortest distances between video frames and clustering centers. The exams show that the extracting key frames can express the main content of video and redundant information is less when without artificial intervention.

Key words: video abstract; manifold learning algorithm; locally linear embedding; adaptive FCM; adaptive threshold FCM

1 引言

随着多媒体技术的快速发展, 视频数据也越来越多. 如何对这些海量视频进行管理, 使用户能快速、准确地检索到他们需要的视频, 是当前信息处理领域的一个研究热点和难点. 视频摘要是基于内容视频检索、视频浏览、视频索引的重要步骤, 它的实

质就是用视频进行场景分割, 以便得到视频场景的代表帧. 视频摘要主要有基于关键帧的视频摘要和基于视频整体内容的视频两类.

由于基于关键帧的视频摘要存在较多问题^[1], 本文研究新的基于关键帧的视频摘要提取算法: LLE^[2]-自适应 FCM 和 LLE-自适应阈值 FCM 算法, 试图消除视频摘要中的信息冗余. 算法的主要步

收稿日期: 2011-07-08; **修回日期:** 2011-08-04

基金项目: 国家自然科学基金项目(61070121, 60973094); 江苏省自然科学基金项目(BK2009538); 江苏省产学研前瞻性联合研究项目(BY2009117)

骤是:首先用流形学习算法 LLE 对视频图像进行非线性降维,然后通过聚类有效性函数(自适应 FCM)或通过阈值的不断自动改变(自适应阈值 FCM)得到分类类别数,将降维后的特征向量和分类类别数用 FCM 算法找到聚类中心,最后计算视频帧与聚类中心的距离,将距离聚类中心最近的视频帧作为关键帧,形成视频摘要。

2 基于 LLE 的视频摘要提取

2.1 LLE 算法

给定 N 个输入向量 $\{x_1; x_2; \dots; x_N\}$, $x_i \in R^d$, 通过 LLE 算法,得到输出向量 $y_i, i \in [1, N]$, $y_i \in R^m$, $m \ll d$. LLE 算法可以归结为三步:(1)寻找每个样本点的 k 个近邻点;(2)由每个样本点的近邻点计算出该样本点的局部重建权值矩阵;(3)由该样本点的局部重建权值矩阵和其近邻点计算出该样本点的输出向量。

算法的第一步是计算出每个样本点的 k 个近邻点,把相对于所求样本点的欧氏距离最近的 k 个样本点规定为所求样本点的 k 个近邻点, k 是一个固定值。

这里定义一个误差函数,用于求取局部重建权值矩阵. 误差函数如下所示:

$$\min \epsilon(W) = \sum_{i=1}^m \left\| X_i - \sum_{j=1}^k W_j^i X_{ij} \right\|^2 \quad (1)$$

其中 $x_{ij} (j=1, 2, \dots, k)$ 为 x_i 的 k 个近邻点, w_j^i 是 x_i 与 x_{ij} 之间的权值,且要满足: $\sum_{j=1}^k w_j^i = 1$. 在求取局部重建权值矩阵 W 时应使误差函数值达到最小,也就是说由样本点的近邻点,构造出最优 W 矩阵使误差函数值达到最小。

最后,将所有的样本点映射为低维空间中的数据,必须要使输出数据在低维空间中保持原有的拓扑结构. 为此构造一个损失函数,映射过程中必须使损失函数值达到最小. 该函数形式如下所示:

$$\epsilon(Y) = \sum_{i=1}^n \left\| Y_i - \sum_{j=1}^k W_j^i Y_{ij} \right\|^2 \quad (2)$$

其中, y_i 是 x_i 的输出向量, $y_{ij} (j=1, 2, \dots, k)$ 是 y_i 的 k 个近邻点,且要满足两个条件,即: $\sum_{i=1}^m y_i = 0$, $\frac{1}{m} \sum_{i=1}^m y_i y_i^T = I$, 其中 I 是 $m \times m$ 的单位矩阵. 然后求取的最优解 y_i 应使损失函数值达到最小. $w_j^i (i=1, 2, \dots, m)$ 可以存储在 $m \times m$ 的稀疏矩阵 W 中,

当 x_j 是 x_i 的近邻点时, $W_{i,j} = w_j^i$, 否则, $W_{i,j} = 0$. 则损失函数可重写为:

$$\min \epsilon(Y) = \sum_{i=1}^m \sum_{j=1}^m M_{i,j} y_i^T y_j \quad (3)$$

其中 M 是一个 $m \times m$ 的对称矩阵,其表达式为:

$$M = (I - W)^T (I - W) \quad (4)$$

由式(4)可知,要使损失函数值达到最小,则取 Y 为 M 的最小非零 m 个特征值所对应的特征向量。

2.2 自适应 FCM

传统的 FCM 算法进行聚类时,需要事先人为的确定聚类数目 c , 因此它只能适用于聚类数目已知的聚类问题. 但对于实际问题(如图像分割), 聚类数目通常是未知的. 如果聚类数选取的不合适, 则会使数据集的划分结果与数据集的真正结构不相符, 可能会将不是一类的数据误分为同一类, 从而导致分类失败. 聚类有效性函数是用于评价聚类结果优劣的指标, 准确地给出初始聚类类别数将使得聚类结果趋于合理化。

本文引入了自适应 FCM 方法^[3], 此方法通过聚类有效性函数^[4]不断调整聚类的数目, 最终得到最佳的聚类数:

$$P_t(U, V; X) = \frac{\sum_{i=1}^c \sum_{j=1}^n (u_{ij})^n \|x_j - v_i\|^2}{n \left[\min_{i,k=1, i \neq k}^c \|v_i - v_k\| \right]} \quad (5)$$

其中 c 为给定的聚类类数, 隶属度矩阵 $U \in M_{fc}$ 是对包含 n 个数据点的数据集 $X = \{x_1, x_2, \dots, x_n\}$ 的一种模糊 c 划分, $V = \{v_1, v_2, \dots, v_c\}$ 为 c 个聚类中心集合。

该方法首先确定可能聚类类别数为 $[c_{\min}, c_{\max}]$, 本文中初始化 $P_t = \inf$, $c_{\min} = 2$, 而 $c_{\max}$ 的值为本文用到的数据集提过的故事板中的帧数作为参照; 然后使用 FCM 算法对数据集进行聚类分析, 同时计算出 $P_t^* < P_t$ 的值, 如果 $P_t^* < P_t$, 则 $P_t = P_t^*$, $c = c^*$, $c = c + 1$; 如果 $c = c_{\max}$ 则停止迭代, 否则继续前面的步骤。

2.3 自适应阈值 FCM

同 2.2 节中提到的自适应 FCM 一样, 自适应阈值也是为了自动获取聚类类别数. 本文在此基础上提出了自适应阈值 FCM, 这种算法相当于二次聚类的方法, 它能够通过不断变化的阈值分割得出聚类的数目, 然后将聚类的数目输入到 FCM 中进行二次聚类, 从而得出数据集的分类和聚类中心. 较自适应 FCM 而言, 此方法完全不需要人工干预, 而

FCM 虽然在一定程度上能够自动得出聚类的类别数,但是需要事先确定 $c_{\min}$ 和 $c_{\max}$ 的大小,如果这两个参数确定的不好,会对分类的效果产生影响。

自适应阈值 FCM 具体算法如下:

步骤 1: 首先假定所有帧都在一个类簇中,而 f_1 就为这个类的聚类中心,而阈值的计算公式为 $T = \left[\frac{1}{N-1} \sum_i^N \text{dist}(f_1, f_i) \right] + c$, 式中 N 为视频的总帧数, c 为常数,族个数 $N_c = 1$;

步骤 2: 计算 f_1 到所有帧的距离 dist , 如果 $\text{dist} < T$ 则将这些帧放在一个类簇中, 如果 $\text{dist} > T$, 将这些放入 theleft 类簇, $N_c = N_c + 1$;

步骤 3: 返回步骤 1, 将类簇的第一帧作为 f_1 , 在 theleft 类簇中重新计算阈值;

步骤 4: 以此往复, 直到循环完成, 最后将聚类的数目送入到 FCM 中从而得到聚类中心。

3 实验结果

为了验证本文算法的效果, 在 matlab 环境下, 从开放视频组织的视频库^[5] 选择了 n 幅视频进行实验, 并将实验的得到的视频摘要与开放视频组织库提供的故事板进行比较。视频组织库中故事板是使用 c-means 方法得到的, 并且经过了人工的干预。本实验主要有以下两个部分: (1) 构造视频的流形表示, 进行降维聚类获取视频摘要; (2) 给出关键帧提取的判别准则。

3.1 提取视频摘要

从开放视频组织的视频库中抽取了 3 个视频, indi008、indi104、apol6005 进行实验。首先对视频数据进行处理, 为了减少计算量, 将视频帧进行 4×4 的分块化处理, 用块的灰度均值代替整个矩阵的灰度, 并将其转换为一维向量; 然后用流形学习算法 LLE 对提取的向量进行降维处理 (经过多次实验将降维的维数确定为 4, 邻域参数的选择视情况而定), 得到特征向量; 将得到的特征向量送入到自适应 FCM/自适应阈值 FCM, 得到聚类中心; 最后将离聚类中心最近的视频帧作为关键帧。

图 1 是通过四种方法提取的 indi008 关键帧。其中 LLE-自适应 FCM 提取的关键帧和视频组织库中的一样, 而 LLE-自适应阈值 FCM 虽然取得的关键帧比 LLE-自适应 FCM 和视频组织库少一帧, 但是却能够反映视频的主要内容。从图 1 中可以看出视频组织库中的第四帧可以是没有的, 因为这一帧如果不存在, 并不影响人们对视频内容的理解。SKF



(a) LLE-自适应 FCM



(b) LLE-自适应阈值 FCM



(c) SKF



(d) 视频组织库

图 1 indi008 视频关键帧

提取的关键帧很明显出现了冗余帧, 由此可以看出 LLE-自适应 FCM 和 LLE-自适应阈值 FCM 提取的关键帧比较适合, 而 LLE-自适应阈值 FCM 的实验结果更精练一些。

对 indi104 视频提取的关键帧表明, LLE-自适应 FCM 提取的关键帧比视频组织库提取的多两帧, 其中第一帧虽然和视频组织库的第一帧主要内

容不一样,但是却是与视频组织库中的第一帧内容相关,而 LLE-自适应 FCM 提取的第二和第三帧在视频组织库中没有,但这并不是冗余帧,这两帧能反映视频的主要内容,这说明视频组织库提供的关键帧因为方法的局限性,所以并不能作为视频的摘要。LLE-自适应阈值 FCM 提取的关键帧基本能反映视频的主要内容,但是却少了两帧。SKF 同样反映的关键帧也不完全。

对 apol6005 视频提取的关键帧表明,当使用 LLE 进行降维的时候选取邻域参数为 35,这样能够尽可能保持数据的主要信息不丢失。通过比较发现,LLE-自适应阈值 FCM 提取的关键帧最能反映视频的主要内容,而其他的几种方法均出现了冗余帧,并且出现漏检现象。当然视频组织库也存在这种现象。

综上所述,在没有人工干涉的情况下,本文提出的算法更能够全面的反映视频的主要内容,而且冗余小,在一定程度上,比自组织视频库中提供的故事版更能够说明问题。通过比较可以看出 LLE-自适应阈值 FCM 较 LLE-自适应 FCM、SKF 的实验效果要好一些,提取出的视频摘要最具代表性。

3.2 关键帧提取准则

衡量关键帧提取算法的性能有两个常用的测试指标:查全率和查准率。检索作为综合查询、描述、匹配、提取的处理,有成功和失败的识别的原理。模式识别原理对应的基本参数,如表 1 所示。

表 2 几种方法的查全率和准确率

视频源	总帧数	LLE-自适应 FCM					LLE-自适应阈值 FCM					SKF				
		提取关 键帧数	漏检 数	误检 数	查全 率	查准 率	提取关 键帧数	漏检 数	误检 数	查全 率	查准 率	提取关 键帧数	漏检 数	误检 数	查全 率	查准 率
indi008	706	5	0	0	1	1	4	1	0	0.8	1	5	1	1	0.83	0.83
indi104	1171	7	0	3	1	0.7	5	2	2	0.71	0.71	5	2	2	0.71	0.71
apol6005	692	5	1	1	0.83	0.83	6	2	2	0.75	0.75	6	2	2	0.75	0.75

4 结束语

视频摘要的提取是当前信息处理领域的一个研究热点和难点,本文将流形学习和聚类方法(自动分类并且得到聚类中心)相结合,能够自动获取关键帧,并且冗余少。本文中用到的算法已取得一定的成效,但是还存在一定的冗余信息,并且算法的稳定性也存在一定欠缺。所以对聚类算法进行改进,以获取更有效的结果,是以后要研究的。

表 1 关键帧提取判别准则

	评价者的判决	
	有关联	无关联
检索到	A(正确检索)	B(误检)
没检索到	C(漏检)	D(正确拒绝)

查全率和准确率的定义如下:

查全率=有关联的正确检索结果/所有有关联的结果= $A/(A+C)$

准确率=有关联的正确检索结果/所有检索到的结果= $A/(A+B)$

查全率是指图像序列中所有的相关对象是否都查到了,查准率是指查到的对象是否都是正确的,均用百分比来表示。

表 2 中的数据主要是基于自组织视频库中提供的故事版而来的,相对于故事版而言,本文提出的算法得出的查全率和查准率不是很理想。但是,自组织视频库中的故事版是基于 C-means 的,而且还有人因素在里面,有的时候提取的关键帧不能够很好的反映视频的主要内容,而本文中 LLE-自适应 FCM 和 LLE-自适应阈值 FCM 在没有人干预下,能够较好的反映视频的主要内容,就算是基于自组织视频库而言,效果也是不错的。相比而言,SKF 得到的结果较差。

参考文献:

- [1] 朱映映,周洞汝.一种从压缩视频流中提取关键帧的方法[J].计算机工程与应用,2003,20(18):13-15.
- [2] Roweis S T. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000,290(5500): 2323-2326.

(下转第 68 页)

5 结束语

本文从基于本体的知识检索系统结构出发,分析构建了一个基于本体知识库的提升机故障知识检索系统.通过原型系统的实现和应用,验证了本文所提出的方法的有效性和正确性,为提升机故障知识模型的检索研究提供了有效的支援和保障.

参考文献:

- [1] 马文峰,杜小勇.知识检索研究[J].情报理论与实践,2006(2):157-161.
- [2] 杨良斌,黄国彬,周静怡.近两年来国外有关本体基本问题的主要研究述评[J].图书馆建设,2008(8):80-83.
- [3] 丁晟春,顾德访. Jena 在实现基于 Ontology 的语义检索中的应用研究[J].现代图书情报技术,2005(10):5-9.
- [4] 许浩.基于本体论的语义检索技术的研究与实现[D].南京:南京理工大学,2007.
- [5] David Vallet, Miriam Fernandez, Pablo Castells. An

ontology-based information retrieval model[C]// Extend Semantic Web Conference. Greece, Aeraklion, 2005: 455-470.

- [6] 聂卉.基于本体的查询扩展与规范[J].现代图书情报技术,2007(3):35-38.
- [7] Xiao Hang Wang, Tao Gu, Da Qing Zhang, et al. Ontology based context modeling and reasoning using OWL[C]// IEEE International Conference on Pervasive Computing and Communication. England: Nottingham, 2004.
- [8] 何琳,曹玲.农业古籍本体的构建及其检索机制研究[J].现代图书情报技术,2006(12):37-40.

作者简介:



许晓林 (1968—),高级工程师.研究方向为大型机电设备技术管理.

(上接第 63 页)

- [3] 宋相法,陈志国.基与 Fuzzy c-means 算法聚类有效性函数的纹理分割[J].河南大学学报:自然科学版,2004,34(1):14-17.
- [4] Xie X L, Beni G. A validity measure for fuzzy clustering [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence,1991, 13(8): 841-847.
- [5] Open_video project[EB/OL]. [2007-03-20] <http://www.open-video.org>.

作者简介:



冯燕女,(1986—),硕士研究生.研究方向为人工智能和图像处理.

王洪元 男,(1960—),博士,教授.研究方向为模式识别与智能系统.

程起才 男,(1981—),硕士,讲师.研究方向为机器学习和数字图像处理.