

DOI: 10.11992/tis.201806013

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20180706.1003.002.html>

基于排序学习的视频摘要

王钊润¹, 聂秀山², 杨帆², 吕鹏², 尹义龙³

(1. 山东大学 计算机科学与技术学院, 山东 济南 250101; 2. 山东财经大学 计算机科学与技术学院, 山东 济南 250014; 3. 山东大学 软件学院, 山东 济南 250101)

摘 要: 视频数据的急剧增加, 给视频的浏览、存储、检索等应用带来一系列问题和挑战, 视频摘要正是解决此类问题的一个有效途径。针对现有视频摘要算法基于约束和经验设置构造目标函数, 并对帧集合进行打分带来的不确定和复杂度高等问题, 提出一个基于排序学习的视频摘要生成方法。该方法把视频摘要的提取等价为视频帧对视频内容表示的相关度排序问题, 利用训练集学习排序函数, 使得排序靠前的是与视频相关度高的帧, 用学到的排序函数对帧打分, 根据分数高低选择关键帧作为视频摘要。另外, 与现有方法相比, 该方法是对帧而非帧集合打分, 计算复杂度显著降低。通过在 TVSum50 数据集上测试, 实验结果证实了该方法的有效性。

关键词: 视频帧; 摘要; 提取视频帧; 排序; 视频操作; 视频图像; 视频; 深度学习

中图分类号: TP389.1 **文献标志码:** A **文章编号:** 1673-4785(2018)06-0921-07

中文引用格式: 王钊润, 聂秀山, 杨帆, 等. 基于排序学习的视频摘要[J]. 智能系统学报, 2018, 13(6): 921-927.

英文引用格式: WANG Xingrun, NIE Xiushan, YANG Fan, et al. Video summarization based on learning to rank[J]. CAAI transactions on intelligent systems, 2018, 13(6): 921-927.

Video summarization based on learning to rank

WANG Xingrun¹, NIE Xiushan², YANG Fan², LYU Peng², YIN Yilong³

(1. School of Computer Science and Technology, Shandong University, Ji'nan 250101, China; 2. School of Computer Science and Technology, Shandong University of Finance and Economics, Ji'nan 250014, China; 3. School of Software Engineering, Shandong University, Ji'nan 250101, China)

Abstract: The exponential increase in the number of online videos has resulted in several challenges as regards video browsing, video storing, and video retrieval. These challenges can be effectively solved by video summarization. The existing video summarization methods construct objective functions based on empirical constraints and experience setup resulting from scoring for a set of frames. However, these methods have uncertainty and high complexity; therefore, in this paper, a video summarization method based on learning-to-rank algorithm is proposed. The proposed method considers summary extraction as a correlation ranking problem between frames and video. First, the training set is used to learn the ranking function, which places the frames having high correlation with video in the front position. Then, the score of each frame is calculated using the learned ranking function. Finally, the keyframes with high scores are selected as the video summary. Compared with the existing methods, the proposed method calculates a score for each frame rather than for a set of frames; therefore, computation complexity remarkably decreases. In addition, the effectiveness of the proposed approach is validated using experimental results on TVSum50 dataset.

Keywords: video frame; summary; video frame grabbers; ranking; video operation; video images; video; deep learning

收稿日期: 2018-06-06. 网络出版日期: 2018-07-06.

基金项目: 国家自然科学基金项目 (61671274, 61573219); 中国博士后基金项目 (2016M592190); 山东省重点研发计划项目 (2017CXGC1504); 山东省高校优势学科人才团队培育计划.

通信作者: 聂秀山. E-mail: nixiushan@163.com.

随着手机、摄像机等录像设备的普及, 视频拍摄越来越简单方便。一项调查显示, 在 YouTube 视频网站, 每天视频的上传时长大约是 14 万小时^[1], 视频数据的爆炸式增长带来了一些不可避免的问题。

题。对于用户来说,浏览14万小时的视频需要不间断地观看大约16年时间,同时,存储如此庞大的视频数据也给网站带来巨大的压力,除此之外,视频检索也要花费更多时间。由于视频数据快速增长带来的一系列问题,视频处理的相关技术也逐渐受到人们的重视。

为了解决由于庞大的视频数据造成的问题,人们提出了视频摘要技术。视频摘要技术是视频处理的一种技术,简单地说,它是从视频中选取几个视频段或者几张图片,被选出来的视频段或图片可以简要概括视频内容。在视频摘要之前,基本上需要花费与视频等长的时间来浏览视频,但是有了视频摘要后,人们只需要观看视频段或图片就可以清楚视频的内容,为浏览视频节省了大量时间。而且,因为视频段或图片基本上包含了视频的主要内容,只要存储视频段或图片即可,为网站视频存储节省了大量空间。同时,在搜索视频时,没必要花费大量的时间搜索整个视频,只需检索相应的视频段或图片。视频摘要技术可以解决视频数据迅猛增长产生的问题,极大地方便了人们的生活。

对于视频摘要的分类,有很多不同的标准。根据输出的摘要类型划分,可以分为动态视频摘要和静态视频摘要^[2]。动态视频摘要是从视频中选取一些视频片段,把这些视频片段组织连贯起来,形成一段流畅的视频作为摘要。静态视频摘要是从视频中选取几帧重要的视频帧,将这些视频帧组织起来构成视频的摘要。这两种形式的视频摘要有各自的特点^[3]。动态视频摘要是一小段视频,包含了音频信息和连续的动作信息,可以帮助用户更加生动地了解视频的主要内容。静态视频摘要是由图片组成的,以时间顺序呈现在用户面前,具有更高的浏览效率。当然,无论是动态视频摘要还是静态视频摘要,都能代表视频内容,都能达到在看过视频摘要后就可以清楚地知道视频内容的效果。

由于视频数据的增多,根据视频内容自动提取视频摘要已经是大势所趋。一般来说,视频摘要都由以下4个步骤组成:特征提取、视频镜头分割、视频内容重要性评价、视频摘要生成^[2]。特征提取是视频处理最基础的一步,提取的视频帧特征有全局特征(例如颜色、纹理、运动信息)和局部特征(例如尺度不变特征变换SIFT)^[4]。近来,也用到了一些高级语义特征和深度特征。视频镜头分割^[2]就是把一长段视频分成几个小片段,满足同一片段内的视频内容尽可能相似,不

同片段间的视频内容尽可能不同,通常是根据视频帧之间的相似度进行视频分段。视频内容重要性评价是指算法按照一些规则和依据,对视频内容的重要性进行评价,为后续提取视频摘要做准备。生成视频摘要是根据重要性评价结果,将其中比较重要的部分提取出来,作为整个视频的摘要,其输出形式可以是动态的短视频或者是静态的图片帧。本文的视频摘要形式是静态的视频帧。

视频摘要的方法有很多,聚类^[5]是主要方法之一,聚类是把相似的视频帧聚成一簇,从每一簇里选取几帧组成视频摘要。之后出现了对视频摘要做约束的方法,例如摘要应该覆盖视频的内容,摘要的冗余性应该比较低等,针对这些约束各自构建了计算公式,然后从视频帧集合中选取一个分数最高的子集作为摘要。然而,这些方法有一些缺点,首先要对摘要做约束,这个是要先验知识的,理解角度不一样对摘要的约束也不同,对摘要的约束个数也不同,约束不同继而构造的计算公式也不一样。还有,摘要是从视频帧集合中选取根据约束公式计算的分数最高的子集。若一个视频有 n 个帧,视频帧集合则有 $2^n - 1$ 种组合,计算复杂度是 $O(2^n)$,每增加一帧计算量会呈指数式增长。针对以上问题,本文提出了一个基于排序学习^[6](learning to rank)的视频摘要方法。

本文工作的主要贡献有以下两点:

1) 本文视频摘要的方法不依赖于先验知识的约束。本文把视频摘要看成是对视频帧的排序,根据训练集训练排序算法,使得和视频相关的帧排在前面的位置。

2) 相比之前的视频摘要方法,本文方法的计算量大大降低。本文是对视频帧打分,计算复杂度是 $O(n)$,之前的大多数方法是对视频帧集合打分,计算复杂度是 $O(2^n)$ 。

本文方法是基于排序学习算法来解决关键帧选取问题,把视频帧选取看成一个排序问题,与视频相关性大的帧被排在前面,这些帧被选为关键帧。本文方法依然是按照视频摘要的4个步骤进行,因为视频的连续性,首先对视频分段,然后提取视频帧的深度特征,之后用排序学习算法对视频中的帧排序,最后选取排在前面的帧组成视频摘要。

1 相关工作

大多数视频摘要的方法基于两个准则,一个是选取的关键帧能尽可能多地包含视频内容,另一个是被选取的关键帧之间尽可能不同。基于这

两个准则,设计了不同的计算公式。Guan 等^[4]提出了基于关键点的关键帧选择 (keypoint-based keyframe selection) 算法,是一种无监督方法,文中给出覆盖率和冗余性两个公式,提取每个视频帧的 SIFT 局部特征,把提取到的所有帧的 SIFT 局部特征组成关键点池,每个视频帧与关键点池进行匹配,从关键点池中去掉已经匹配的关键点,能最多覆盖关键点池并且可以最小化摘要冗余性的帧被选为关键帧。Chakraborty 等^[6]设计了代表性和独特性两个公式,代表性是度量关键帧集合与视频的相似性,独特性是量化关键帧集合中帧之间的相似性,赋予代表性和独特性合适的权重来计算候选集合的得分,得分最高的集合被选为关键帧集合。Gong 等^[1]提出了 seqDPP(sequential determinantal point process) 模型,它是一个概率模型,是基于 DPP 模型做了改进。DPP 模型可以确保选择的关键帧之间互不相同,但是却没有考虑到视频的时序性。例如,一个视频中开始部分包含了吃早餐的镜头,结束部分包含了吃晚餐的镜头。如果使用 DPP 模型它只会从早餐和晚餐里选一个镜头,但是因为吃早餐和吃晚餐是两件不同的事情,而且相隔时间比较长,所以这两个镜头应该都被选为关键帧。为了弥补这个缺陷,Gong 等先把视频分割成几个小片段,在每个小片段里使用 DPP 算法,在当前片段里选取关键帧时要考虑到前一片段中已经选取的关键帧,避免当前片段选取的关键帧与前一片段选取的关键帧过于相似。Li 等^[5]提出了 4 个模型,分别是重要性、代表性、多样性和故事性,重要性是指选取的关键帧要包含重要的人和物,代表性是指选取的关键帧能代表视频内容,多样性是指选取的关键帧要尽可能不同,故事性是指选取的关键帧故事性比较强,用户能比较容易理解视频内容。相应地构建了 4 个公式并且学习得到关于 4 个模型的权重,最终用于计算关键帧集合的分数,分数最高的关键帧集合被选为视频摘要。Hu 等^[7]提出了用多个属性和图片质量来提取摘要,在文中提出了 9 个属性和一个计算图片质量的方法。Hu 等认为摘要中的帧应该是清楚的、清晰的,而视频中的一些帧质量不是高的,有可能是失真的、模糊的,这些图片不应该被选择到摘要中,所以计算了每帧的质量作为这一帧可以被选为关键帧的权重。Sun 等^[8]提出了 SASUM(semantic attribute assisted video SUMmarization) 的视频摘要方法,学习了一个深度神经网络用来提取每一帧的语义特征,用得到的语义特征把图片聚成几组,选取每组的中心片段组成最终的视频摘要。

大多数的视频摘要方法都是根据个人经验构建模型,然后学习各个模型的权重,之后计算视频帧集合的分数,从中选取分数最高的视频帧集合作为视频摘要。对于这类方法,首先要构建模型,模型设计的好坏,模型的个数对摘要的结果有很大的影响。除此之外,还要计算所有视频帧集合的分数,一个仅有 20 个帧的视频就会产生一百多万种组合,计算复杂度是 $O(2^n)$ 。

本文的视频摘要方法是对视频帧打分,使得自动产生的分数分布与人工标记的分数分布尽可能吻合。学习打分的过程是基于排序学习算法,该算法不仅考虑到帧与视频的关系,也考虑到帧与帧之间的关系。本文的方法直接利用人工标记的摘要训练学习器,不依赖先验知识的约束。而且,本文的方法是对视频帧打分,计算复杂度是 $O(n)$,相比于对帧集合打分的方法,复杂度大大降低。

2 基于排序学习的视频摘要

2.1 排序学习

排序学习被广泛应用于文档检索,给予一个查询条件,排序学习算法会给出与查询条件相关的文档关于相关性的一个由高到低的排序,排在越前面的文档是越符合查询条件的文档。排序学习算法有基于点、基于文档对^[9]、基于文档列表 3 种方法。视频摘要与基于文档列表排序的思想更切合,在视频摘要中借鉴的是基于文档列表的方法。

Listwise 方法对文档的排序结果进行优化,使得预测的排序与 ground truth 的排序更接近。在训练阶段,把排序函数自动产生的分数与 ground truth 中的分数转换成概率,衡量两个概率分布的误差,误差越小说明自动排序结果和 ground truth 排序结果越接近。预测时直接用训练好的排序函数对文档打分,分数越高说明与查询条件越相关。

在 Listwise 方法里有两种概率模型,分别是 Permutation Probability 和 Top k Probability,概率模型把文档得分转换成概率,计算所有排列的概率组成概率列表。构建的概率模型应该满足与 ground truth 分布越接近的分布发生的概率越大。每种排列对应一个概率,所有可能的排列的概率之和为 1。Permutation Probability 考虑了所有文档的排列,Top k Probability 仅考虑了 k 个文档的排列。本文采用的是 Top k Probability,用公式表示为

$$P_s(\zeta_k(j_1, j_2, \dots, j_k)) = \prod_{i=1}^k \frac{\varphi(s_{j_i})}{\sum_{l=i}^n \varphi(s_{j_l})} \quad (1)$$

式中: ζ_k 表示 n 个文档中 k 个文档的排列,总共有

$\frac{n!}{(n-k)!}$ 种情况; j_i 表示排列中第 i 个文档; s_{j_i} 表示文档 j_i 对应的分数。

2.2 排序学习视频摘要方法详述

视频摘要要是把与视频相关的视频帧按照相关性排序并呈现给用户,不同的视频时长不一样。文档检索是根据文档与查询条件的相关性由高到低排序,把排序后的文档呈现给用户,不同的检索条件相关的文档数目不一样。文档检索是要学习文档的顺序,自动产生的相关性高的文档要尽可能与人工标注的一样,视频摘要要是学习视频帧的顺序,自动产生的与视频相关性高的帧要尽可能与人工标注的一样。在这里视频相当于查询条件,视频帧相当于文档。基于排序学习的文档检索的广泛应用,证明排序学习可以很好地学习到人工排序的过程,视频摘要和文档检索是如此相似,因此本文提出了基于排序学习的视频摘要,把视频摘要建模成视频帧与视频的相关性排序。主要分为以下5个步骤:

1) 视频预处理:由于视频的连续性,不能直接对视频使用排序学习算法。如果直接使用排序学习算法会导致选出来的摘要中有太多相似的图片,这与摘要中的帧应尽可能不同相悖。所以,在排序学习算法之前需要对视频分段,在本文中是把视频2s分为一段。

2) 特征提取:本文用到的特征是视频帧的深度特征^[10],是用预训练的VGG-19卷积神经网络提取的。VGG-19网络包含了16个卷积层和3个全连接层。每个视频帧用预训练的VGG-19卷积神经网络提取到4096维特征,然后用PCA算法对4096维特征降维,属于同一个视频的帧的特征平均后就是该视频的特征。

3) 学习排序函数:在对视频每一帧打分之前,首先要根据训练样本学习排序函数,有了排序函数后再对视频中的每一帧打分。

在本文的符号表示中,上角标表示视频帧id,下角标表示视频id。视频集合 $Q = \{q^{(1)}, q^{(2)}, \dots, q^{(m)}\}$, 视频 $q^{(i)}$ 的视频帧列表 $d^{(i)} = (d_1^{(i)}, d_2^{(i)}, \dots, d_{n^{(i)}}^{(i)})$, $d_{n^{(i)}}^{(i)}$ 表示视频 $q^{(i)}$ 第 $n^{(i)}$ 个视频帧。每个视频帧列表 $d^{(i)}$ 都有一组相关性分数 $y^{(i)} = (y_1^{(i)}, y_2^{(i)}, \dots, y_{n^{(i)}}^{(i)})$, $y_{n^{(i)}}^{(i)}$ 表示视频帧 $d_{n^{(i)}}^{(i)}$ 与视频 $q^{(i)}$ 的相关程度的得分。

对于视频-帧对 $(q^{(i)}, d_j^{(i)})$, $i = 1, 2, \dots, m$; $j = 1, 2, \dots, n^{(i)}$, 它的特征向量是 $\mathbf{x}_j^{(i)} = \psi(q^{(i)}, d_j^{(i)})$ 。特征向量列表 $\mathbf{x}^{(i)} = (\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \dots, \mathbf{x}_{n^{(i)}}^{(i)})$, 相应的分数列表 $y^{(i)} = (y_1^{(i)}, y_2^{(i)}, \dots, y_{n^{(i)}}^{(i)})$ 。 $\mathbf{x}^{(i)}$ 和 $y^{(i)}$ 构成一个训练实例,训练集合表示成 $\tau = \{(\mathbf{x}^{(i)}, y^{(i)})\}_{i=1}^m$ 。

构建一个排序函数 f , 对于每一个特征向量

$\mathbf{x}_j^{(i)}$ 都可以得到一个分数 $f(\mathbf{x}_j^{(i)})$, 对于特征向量列表 $\mathbf{x}^{(i)}$ 可以得到分数列表 $z^{(i)} = (f(\mathbf{x}_1^{(i)}), \dots, f(\mathbf{x}_{n^{(i)}}^{(i)}))$ 。目标函数可以形式化为 $\sum_{i=1}^m L(y^{(i)}, z^{(i)})$, 其中 L 是损失函数,只需要使得目标函数的误差最小即可。

对于视频摘要来说,没必要使得预测的概率列表和 ground truth 的概率列表,相应位置对应的排列中所有视频帧的顺序相同,只要使得分数第一高的帧排在该排列的第一个位置,分数第二高的帧排在排列第2个位置,以此类推,即只要排列中的第1个帧相同即可,即排序学习算法 Top k Probability 中的 $k=1$ 。还有,本文的排序函数 f 中特征和权重是线性关系, $\varphi(\cdot)$ 是指数函数。

用交叉熵损失函数衡量视频 $q^{(i)}$ 的损失为

$$L(y^{(i)}, z^{(i)}(f_\omega)) = - \sum_{g \in \zeta_k} P_{y^{(i)}}(g) \log(P_{z^{(i)}(f_\omega)}(g)) \quad (2)$$

$k=1$, 式(1)可以表示成:

$$P_{z^{(i)}(f_\omega)}(\zeta_k(j_1, j_2, \dots, j_k)) = \prod_{i=1}^k \frac{\exp(f_\omega(\mathbf{x}_{j_i}^{(i)}))}{\sum_{l=1}^{n^{(i)}} \exp(f_\omega(\mathbf{x}_{j_l}^{(i)}))} \quad (3)$$

由式(3)得

$$\log(P_{z^{(i)}(f_\omega)}(\mathbf{x}_j^{(i)})) = f_\omega(\mathbf{x}_j^{(i)}) - \log\left(\sum_{j=1}^{n^{(i)}} \exp(f_\omega(\mathbf{x}_j^{(i)}))\right) \quad (4)$$

把式(4)代入式(2)得

$$\Delta\omega = \frac{\partial L(y^{(i)}, z^{(i)}(f_\omega))}{\partial \omega} = - \sum_{j=1}^{n^{(i)}} P_{y^{(i)}}(\mathbf{x}_j^{(i)}) \frac{\partial f_\omega(\mathbf{x}_j^{(i)})}{\partial \omega} + \frac{1}{\sum_{j=1}^{n^{(i)}} \exp(f_\omega(\mathbf{x}_j^{(i)}))} \sum_{j=1}^{n^{(i)}} \exp(f_\omega(\mathbf{x}_j^{(i)})) \frac{\partial f_\omega(\mathbf{x}_j^{(i)})}{\partial \omega} \quad (5)$$

之后用梯度下降法求解。

4) 打分:前面根据训练样本学习到的排序函数被认为是学到了人选取关键帧的一个过程,现在可以利用学习到的排序函数

$$P_{Z^{(i)}(f_\omega)}(\mathbf{x}_j^{(i)}) = \frac{\exp(f_\omega(\mathbf{x}_j^{(i)}))}{\sum_{j=1}^{n^{(i)}} \exp(f_\omega(\mathbf{x}_j^{(i)}))} \quad (6)$$

对视频中的帧打分。

5) 提取关键帧:本文生成摘要的方法是把视频帧看成一个排序问题,和视频相关性大的帧排在前面。利用前面学到排序函数对视频中的帧打分,得分高的帧就被选为关键帧,计算复杂度是 $O(n)$ 。大多数视频摘要方法,把生成摘要看成一个优化的问题,从所有的视频帧集合中找出一个得分最高的集合作为视频摘要,计算复杂度是 $O(2^n)$ 。

3 实验与结果

3.1 实验设置

3.1.1 数据集

实验用的数据集是 TVSum50 数据库^[11], 包含

50 个视频, 10 个类别, 每个类别有 5 个视频, 视频时长 2~10 min 不等, 视频包含了新闻、纪录片、用户拍摄等不同的种类, 视频被每 2 s 分成一段, 每个视频段由 20 个用户打分, 产生 20 个分数 (1~5, 5 代表该视频段与视频最相关, 依次递减), 分数高

的视频段中的帧被选为关键帧。就像其他论文中那样, 摘要的长度被限制在小于视频长度的 15%。从每个类别的视频里随机选取一个视频作为测试视频, 剩下的视频作为训练视频, 也就是 40 个视频作为训练集, 10 个视频作为测试集。如图 1 所示。

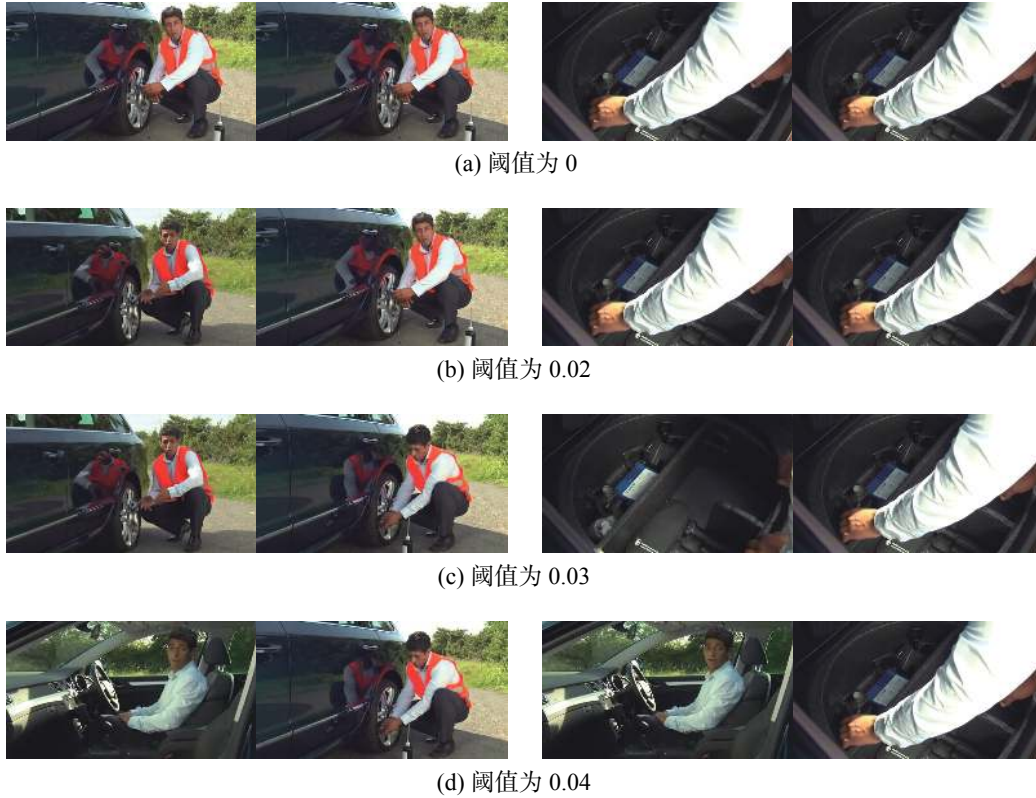


图 1 判断两个图片相似的情况, 从每个阈值里选取了两组图

Fig. 1 Judging the similarity between the two pictures and selecting two pairs of pictures from each threshold

3.1.2 评估方法

对于自动产生的摘要 A 和人工标记的 ground truth 摘要 B , 本文是通过计算摘要 A 和摘要 B 的匹配程度来判断摘要 A 的好坏。从摘要 A 和摘要 B 中分别取出一个视频帧组成一个包含两个视频帧的对, 若 A 中有 n_A 个帧, B 中有 n_B 个帧, 则会产生 $m_A \times n_B$ 种对。计算每对的距离, 若距离小于某个阈值, 则认为这个视频帧对成功匹配, 需要注意的是每个视频帧只能成功匹配一次。精确率 P_{AB} 、召回率 R_{AB} 和 F -score 定义为

$$P_{AB} = \frac{m}{n_A}, R_{AB} = \frac{m}{n_B} \quad (7)$$

$$F\text{-score} = \frac{(1 + \beta^2) \times P_{AB} \times R_{AB}}{(\beta^2 \times P_{AB}) + R_{AB}} \quad (8)$$

其中 m 表示视频帧对匹配成功的个数, β 用来决定精确率和召回率之间的重要程度, 本文设置 $\beta = 1$ 。

用自动产生的摘要分别与 20 个人工标记的摘要计算 P 、 R 、 F -score, 之后取平均值作为该视

频的最终结果。

3.2 实验和结果

3.2.1 阈值设置

视频帧用不同的特征提取方法得到的帧是不一样的, 判断两个视频帧是否相似的阈值也随之不同, 若阈值设置太大会产生两个完全不同的图片被判断为相似, 使得 F -score 值偏大; 反之, 若阈值设置太小会使得 F -score 值偏小, 所以要判断阈值为为何值时认为两个帧相似会比较合理。

从图 2 中可以看出 F -score 随阈值的增大而增加。设置阈值是为了判断两个图片是否相似, 阈值设置太大就不能解决这个问题, 所以要选取合适的阈值。

从图 2 中可知, threshold=0 时, 只有两个图片一模一样才会被判断相似; threshold=0.04 时, 观察发现两个不相似的图片被认定相似, 也就是说阈值设置太大了; 最终 threshold 设置为 0.03。

3.2.2 实验改进

uniform sample 和 random sample 都是取样中

的方法, uniform sample 是按照固定间隔抽取视频帧, random sample 是随机抽取关键帧。聚类是从较大的簇中选取视频帧, 用了两种聚类方法 k -均值聚类 (k -means cluster) 和谱聚类 (spectral cluster)。LiveLight (online video highlighting) 方法^[12]是通过字典衡量冗余信息, 删除冗余信息来选取摘要。sample、cluster 和 LiveLight 都是非监督方法, 而本文的方法是有监督的方法, 用人工标记的摘要学习排序函数, 学习人工打分的过程, 实验结果表明本文的方法更好, 如表 1 所示。

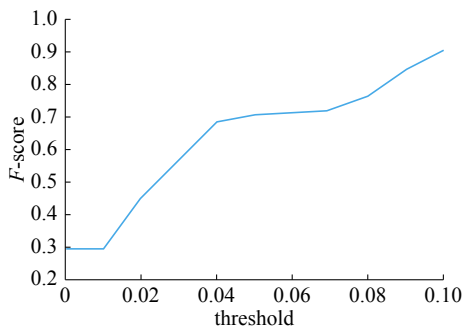


图 2 阈值取不同值时对应的 F -score 值

Fig. 2 F -score value with different threshold

表 1 不同方法在 TVSum50 上的实验结果

Table 1 The Results on TVSum50

方法	F -score
uniform sample	0.36
random sample	0.32
k -means cluster	0.35
spectral cluster	0.39
LiveLight	0.46
Our	0.48

4 结束语

由于目前大多数的视频摘要方法要对摘要做约束并构建相应的公式, 而且还要从众多的视频帧集合中挑选比较好的集合作为摘要, 不仅需要先验知识, 由于集合数目太多, 还会增加计算量。本文呈现的视频摘要的方法, 把提取摘要看作是对视频帧的排序问题, 利用人工标记的关键帧训练排序函数, 使得排在前面的帧是人工标记的帧, 因此不是对视频帧集合打分, 而是对单个帧打分, 这样可以减少计算量, 在 TVSum50 数据集上表现出比其他几个视频摘要的方法好。

参考文献:

- [1] GONG Boqing, CHAO Weilun, GRAUMAN K, et al. Diverse sequential subset selection for supervised video summarization[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada, 2014: 2069–2077.
- [2] 李佳桐. 自适应视频摘要算法研究[D]. 合肥: 中国科学技术大学, 2017.
LI Jiatong. Research on adaptive video summarization algorithms[D]. Hefei: University of Science and Technology of China, 2017.
- [3] FURINI M, GERACI F, MONTANGERO M, et al. STIMO: STill and MOving video storyboard for the web scenario[J]. Multimedia tools and applications, 2010, 46(1): 47–69.
- [4] GUAN Genliang, WANG Zhiyong, LU Shiyang, et al. Keypoint-based keyframe selection[J]. IEEE transactions on circuits and systems for video technology, 2013, 23(4): 729–734.
- [5] LI Xuelong, ZHAO Bin, LU Xiaoqiang. A general framework for edited video and raw video summarization[J]. IEEE transactions on image processing, 2017, 26(8): 3652–3664.
- [6] CHAKRABORTY S, TICKOO O, IYER R. Adaptive key-frame selection for video summarization[C]//Proceedings of 2015 IEEE Winter Conference on Applications of Computer Vision. Waikoloa, USA, 2015: 702–709.
- [7] HU Tongling, LI Zechao, SU Weiyang, et al. Unsupervised video summaries using multiple features and image quality[C]//Proceedings of 2017 IEEE Third International Conference on Multimedia Big Data. Laguna Hills, USA, 2017: 117–120.
- [8] SUN Ke, ZHU Jiasong, LEI Zhuo, et al. Learning deep semantic attributes for user video summarization[C]//Proceedings of 2017 IEEE International Conference on Multimedia and Expo. Hong Kong, China, 2017: 643–648.
- [9] CAO Yunbo, XU Jun, LIU Tieyan, et al. Adapting ranking SVM to document retrieval[C]//Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Seattle, Washington, USA, 2006: 186–193.
- [10] 毕晓君, 冯雪赟. 基于改进深度学习模型 C-GRBM 的人体行为识别[J]. 哈尔滨工程大学学报, 2018, 39(1): 156–162.
BI Xiaojun, FENG Xueyun. Human action recognition based on improved depth learning model C-GRBM[J]. Journal of Harbin engineering university, 2018, 39(1): 156–162.
- [11] SONG Yale, VALLMITJANA J, STENT A, et al. Tvsum:

summarizing web videos using titles[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 5179–5187.

- [12] ZHAO Bin, XING E P. Quasi real-time summarization for consumer videos[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 2513–2520.

作者简介:

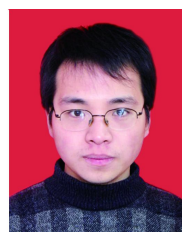


王 鉞 润, 女, 1994 年生, 主要研究方向为机器学习、多媒体信息处理。



1 项、青年项目 1 项, 发表学术论文 30 余篇。

聂秀山, 男, 1981 年生, 教授, 博士, 主要研究方向为机器学习、多媒体信息处理。中国计算机学会人工智能与模式识别专委会委员、中国人工智能学会机器学习专委会通讯委员, 中国计算机学会计算机视觉专委会委员。主持国家自然科学基金面上项目



杨帆, 男, 1983 年生, 主要研究方向为机器学习、凸优化、生物医学。

第三届信息科学、计算机技术与交通运输国际学术会议 The 2018 3rd International Conference on Information Science, Computer Technology and Transportation (ISCTT 2018)

一、会议简介

第三届信息科学、计算机技术与交通运输国际学术会议 (ISCTT2018) 将于 2018 年 12 月 28—30 日在中国西安召开。ISCTT 2018 将围绕“信息科学”、“计算机技术”与“交通运输”的最新研究领域, 为来自国内外高等院校、科学研究所、企事业单位的专家、教授、学者、工程师等提供一个分享专业经验, 扩大专业网络, 面对面交流新思想以及展示研究成果的国际平台, 探讨本领域发展所面临的关键性挑战问题和研究方向, 以期推动该领域理论、技术在高校和企业的发展和应用, 也为参会者建立业务或研究上的联系以及寻找未来事业上的全球合作伙伴。

第三届信息科学、计算机技术与交通运输国际学术会议 (ISCTT 2018) 诚邀领域内专家学者参会!

二、论文评审及出版

1. All accepted full papers will be published by *Journal of Physics: Conference Series (JPCS)* and will be submitted to **EI Compendex, Scopus and CPCI for indexing**. ISCTT2018 会议的论文将被 EI 目录系列期刊 *Journal of Physics: Conference Series (JPCS)* 出版, 出版后提交 EI 和 CPCI 检索。

2. Selected 10 papers will be recommended for publication in SCI journal. 推荐 10 篇优秀论文到 SCI 期刊, 录满截止, 欢迎投稿!

重要信息:

大会官网: www.keoaeic.org/ISCTT2018;

大会时间: 2018 年 12 月 28—30 日;

大会地点: 中国, 西安;

收录检索: EI、CPCI & SCI。