文章编号: 1001- 2486(2009) 06- 0121- 05

新闻视频中基于“场景词汇”的故事单元相似度分析*

文 军[1,2], 吴玲达[1], 曾 璞[1], 谢毓湘[1]
(1. 国防科技大学 信息系统与管理学院, 湖南 长沙 410073; 2. 国防科技大学 理学院, 湖南 长沙 410073)

摘 要: 新闻视频中故事单元的相似度计算对于视频浏览、检索和跟踪故事单元等应用具有特别重要的意义。研究提出了一种利用“场景词汇”计算故事单元相似度的方法, 将单独的关键帧作为一个完整“词汇”, 将每个故事单元看作是一系列“场景词汇”描述的“文档”。在此基础上研究了“场景词汇”的特殊性, 并设计了有效的“场景词汇”权重计算方法和故事单元相似度计算方法。实验显示, 基于“场景词汇”的故事单元相似度计算方法能够比较好地贴近用户感官和实际应用需求。

关键词: 新闻视频; 故事单元; 相似度; 场景词汇
中图分类号: O23 文献标识码: A

Analyzing Similarity of News Video Stories Based on “Scene Words”

WEN Jun, WU Ling-Da, ZENG Pu, Xie Yu-xiang
(1. College of Information Systems and Management, National Univ. of Defense Technology, Changsha 410073, China;
(2. College of Science, National Univ. of Defense Technology, Changsha 410073, China)

Abstract: Analyzing similarity of news video stories is crucial for searching, retrieving, browsing and tracking news video stories. A method based on “scene words” is proposed for measuring the similarity of news stories. Each key frame is seen as a “word” and a news story is seen as a document which is represented by a sequence of “scene words”. In light of this, the particularity of “scene words” is investigated. An effective method to describe weight of words and calculating similarity of stories is presented. Experiment results show that the method based on “scene words” is closer to the sensory and requirement of users.

Key words: news video; story; similarity; scene words

不同来源和不同语种的新闻视频数量很多, 为了有效地检索、浏览和跟踪多语种的新闻视频故事单元, 对故事单元进行相似度计算具有关键性作用。之前, 部分研究利用自动语音识别所得的文本信息进行新闻视频故事单元相似度分析、线程化组织[1-3], 然而因为不同来源视频在语种上的差异, 导致自动语义识别和机器翻译等技术效果受限, 难以获得有效的文本信息, 因而这种方法有很大的局限性。Wu 等[4] 利用文本信息和视觉概念来衡量故事单元之间的相似度, 然而这种方法也面临文本信息和视觉概念获取的困难。一部分视频检索系统利用关键帧底层视觉特征来衡量故事单元之间的相似程度[5-6], 然而在实际研究中, 虽然不同电视台在同一时间报道相同事件的视频包含有相似的人物和场景, 但是因为视频拍摄角度、光照条件、相机传感器或者视频编辑手段有所变化, 导致视觉上存在一定的差异, 这一特点使得基于颜色等底层全局特征来计算故事单元相似度的方法在多源新闻视频数据库中效率低下。因此对不同来源的新闻视频计算故事单元相似度仍然是一个挑战性问题。

1 相关研究

新闻视频故事单元相似度分析可以有效辅助新闻视频检索、新闻事件主题探测和跟踪、数据库组织等各种服务需求, 因此国内外在相关领域开展了大量的研究[11]。研究了同一天中两个不同的英语电视台之间新闻的联系问题, 将关键帧分类为人物关键帧和不包含人物的关键帧分别进行相似度计算, 计算

* 收稿日期: 2009- 03- 25
基金项目: 国家 863 计划资助项目(2009AA01Z335); 国家科技支撑计划资助项目(2007BAH14B01)
作者简介: 文军(1976—), 男, 讲师, 博士。

中融合了部分文本信息。其他初期的相关研究[13-14]也主要通过底层的全局特征(例如:HSV 颜色直方图)来衡量图像相似度[15]。使用颜色直方图来衡量视觉相似性,然后使用时间距离来扩展这种相似性。然而基于全局特征的分析方法对于分析不同来源、不同时间故事单元之间的镜头相似性易于受到光照、编辑方式等各种因素的干扰,使得建立在全局特征基础之上的镜头相似度衡量方法不够鲁棒[11]。

随着图像处理研究中局部特征提取与分析技术的进步,最近的研究证明了使用基于局部关键点和局部特征来克服几何与光学变化对图像匹配的影响是有效的[16]。因此近年来国际研究开始聚焦于基于局部特征,利用相似关键帧识别的信息进行故事单元相似关系分析。相似关键帧(Near-Duplicate Keyframes,NDK)是指一组描述相同或者相似场景但是在视觉上有一定区别的关键帧[7],如图 1 所示,图(a)是不同电视台在同一场景的拍摄,在拍摄时间上的不同导致视觉上有差异;图(b)是素材重复利用时采用不同编辑手段导致视觉上出现差异。NDK 为新闻视频故事单元相似度计算和主题线程化提供了一种有效线索[8-10]。

图 1 相似关键帧的示例

Fig. 1 Samples of near-duplicate keyframes

然而,大部分利用 NDK 进行故事单元相似关系分析的方法在当前研究和应用中都存在局限:绝大部分研究只是简单地利用故事单元之间出现 NDK 就判断为语义相似的,而缺乏对 NDK 信息的相似关系进行定量分析与计算。很多情况下,故事单元之间的相似度仍然借助于关键帧底层全局视觉特征来进行计算。

针对这种情况,本文研究提出了视频关键帧“场景词汇”的概念。“场景词汇”是将故事单元中每个镜头关键帧看作是一个单独的特殊“词汇”,NDK 被看作是相同的词汇,故事单元被转换为由这些“场景词汇”构成的特殊类型“文档”,如图 2 所示。图 2 中两个故事单元分别来源于 CNN 和 BBC,故事单元内部和故事单元之间的 NDK 及其表示的相似关系分别用不同的连线表示,故事单元 1 包含有 6 个场景词汇,其中 W_{13} 和 W_{14} 重复出现,词汇出现频率为 2;故事单元 2 也包含 6 个场景词汇,其中有 3 个词汇在故事单元 1 中也出现,例如:故事单元 2 中的 W_{21} 与故事单元 1 中的 W_{12} 为相同“场景词汇”。

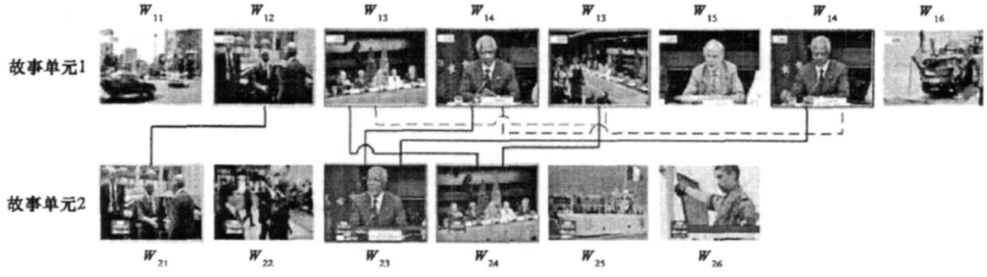

图 2 故事单元场景相似度示意图

Fig. 2 Similarity of news stories based on scenes

与文本中的词汇和底层视觉特征中的“视觉词汇”相比,“场景词汇”具有一定的特殊性:

(1)每个故事单元中的“场景词汇”数量很少,且都为有效词汇;

(2)“场景词汇”包含有更多的信息,一个“场景词汇”能够反映多个“视觉词汇”和“文本词汇”;

(3)故事单元之间相同“场景词汇”的数量很少,大部分“场景词汇”只在一个故事单元中出现,一个故事单元中大部分“场景词汇”的频率为 1;

(4) 故事单元之间出现相似关键帧则预示着具有很强的语义相似性, 因此故事单元之间出现相同“场景词汇”对于故事单元相似程度具有很强的信息支持。

由上述分析可以看出,“场景词汇”与信息检索中的其他类型词汇相比具有特殊属性, 如果直接利用当前信息检索领域的计算方法不能够很好地体现“场景词汇”的特定属性, 尤其是研究要考虑故事单元之间共有的“场景词汇”对故事单元语义相似度的信息支持问题。

2　基于“场景词汇”的故事单元相似度计算方法研究

本文研究在相似关键帧识别的基础上展开, 相似关键帧识别利用我们已经研究的精简局部关键点和SIFT 特征描述子匹配的层次化过滤分析方法[12]。

在信息检索研究中, 进行信息描述和相似度计算最常用的模型有向量空间模型和概率模型, 而概率模型的典型代表是语言模型。

以图 2 中的两个故事单元为例, 二者各有 6 个“场景词汇”, 其中有 3 个是共有的“场景词汇”, 因此这两个故事单元之间具有极高的语义相似度, 其中共有的“场景词汇”提供了强有力的支持。以向量空间模型为例: 故事单元表示为 $S(Kw_1, Kw_2, \cdots, Kw_n)$, 其中 Kw_i 表示“场景词汇”。如果按照常规的TF-IDF 权重计算方法, 认为文档之间的共有词汇分类能力有限, 故而赋予较低的阈值, 这与“场景词汇”的实际情况刚好相反, 例如: 故事单元 1 中“场景词汇” W_{11}的权重约为 0.5636, 而 W_{13} 的权重约为 0.1375, 这与二者对故事单元相似度的信息贡献完全不符(上述计算中对数底数设为 2, 并进行归一化处理)。而两个故事单元之间相似度通过上述计算的结果约为 0.0266, 这与用户主观感官的实际情况完全相反。

而以一元语言模型为例: 图 2 中两个故事单元之间大部分场景词汇不相同, 因此计算的特征项概率为 0, 需要进行平滑处理; 另一方面, 故事单元中“场景词汇”的权重只是统计词汇在各个故事单元中的概率来进行似然性计算, 没有充分考虑共有“场景词汇”对于语义相似度的支持程度。因此, 虽然当前部分研究直接利用了文本领域中的一元语言模型算法, 并没有考虑“场景词汇”的特殊性。并且, 在本质上一元语言模型与向量空间模型都假设词汇之间相互独立, 具有一定相似性, 只是权重计算方法和相似度计算方法上有所差异。

本文研究针对两种模型的特点, 以向量空间模型为基础, 结合概率模型的部分思想, 提出了一种改进的“场景词汇”的权重计算方法, 计算公式如下:

$$t_i = tf(w_i) \times idf(w_i) \Big/ \sqrt{\sum_{i=1}^{m} [tf(w_i) \times idf(w_i)]^2}$$

式中, $tf(w_i)$ 为词汇 w_i 在该故事单元内部出现的频率, m 为该故事单元中词汇的数量, $idf(w_i)$ 为出现词汇 w_i 的故事单元的数量。

上述计算公式融合了词汇在故事单元中的出现概率信息、在故事单元间的“文档频率”信息, 更加突出了故事单元之间共有“场景词汇”的信息贡献。与一元语言模型等概率模型的权重计算相比, 突出了故事单元之间共有“场景词汇”文档频率的信息贡献, 与向量空间模型中TF-IDF 等常用权重计算方法相比, 突出了词汇出现概率和共有“场景词汇”的信息贡献。

按照文本权重计算方法, 仍然以图 2 中故事单元 1 为例,“场景词汇” W_{11} 的权重约为 0.1601, 而 W_{13} 的权重约为 0.6405, 显然这种权重信息更加能够体现二者对于故事单元语义相似性的支持程度。通过本文方法计算的两个故事单元之间相似度为 0.8269, 比较好地贴近了用户的主观感官和实际情况。

在上述算法基础上, 对于故事单元之间的场景相似度计算方法如下:

首先将故事单元表示为 $S(Kw_1, Kw_2, \cdots Kw_n)$, 其中 Kw_i 为故事单元中的“关键帧词汇”, 研究中的算法如下:

步骤 1　对需要计算场景相似度的新闻故事单元 $S(Kw_1, Kw_2, \cdots Kw_n)$, 结合相似关键帧识别的结果, 进行“场景词汇”统计, 对词汇向量进行量化, 得到故事单元的量化表示: $S\{(Kw_1, f_1), (Kw_2, f_2), \cdots, (Kw_n, f_m)\}$, 其中, Kw_i 表示词汇, f_i 表示关键帧在故事单元中出现的频率, 且 $m \leqslant n$;

步骤 2 将需要进行相似度计算的故事单元利用本文提出的方法计算每个关键帧“场景词汇”的权重,将故事单元量化为 $S\{(Kw_1, t_1),(Kw_2, t_2), \cdots,(Kw_n, t_m)\}$,其中 t_i 为关键词的权重;

步骤 3 利用余弦距离计算两个故事单元之间场景相似度:

$$R_f = \frac{\sum_{i=1}^{m}(t_{1,i}t_{2,i})}{\sqrt{\sum_{i=1}^{m}t_{1,i}^2\sum_{j=1}^{m}t_{2,j}^2}}$$

式中,m 为两个故事单元的“场景词汇”总数。

如果故事单元中只有播音员的场景,没有出现其他事件内容描述的场景,则定义 $R_f = 0$。

3 实 验

对于报道相同新闻事件的新闻视频故事单元,不同的用户对于故事单元之间的相似性有不同的认识和要求,具有较强的主观性,难以通过一种标准的方法进行实验,因此实验主要以用户调查的方式来开展。

实验的素材是动态采集的新闻视频数据,采集的节目来源包括 CCTV、BBC 和 CNN,延续时间为 2 周,调查用户重点关注的 3 个事件:“伊朗核问题”、“联合国在黎巴嫩的维和行动”和“伊拉克爆炸事件”,共包含有 45 个不同来源的故事单元。实验选择 9 名没有经过任何训练的用户,将其分成 3 组,分别对故事单元相似度判断进行评估。故事单元相似度判断的评价方式是系统为每个故事单元选择一个最相似的故事单元,将结果提交用户进行判断。在实验中将新闻视频中的故事单元以镜头关键帧的故事板形式提供给用户,通过用户对算法选择信息的判断吻合程度进行比较,计算百分比(将用户对系统计算结果评判为正确的数量与故事单元数量相比)。此外,实验对本文方法与基于单独关键帧全局颜色特征的故事单元相似度计算方法在用户判断一致性上进行比较,实验结果如表 1 所示。

表 1 相似度评估结果

Tab. 1 Evaluating of coherence about similarity

	第一组		第二组		第三组	
故事单元相似程度判断的一致性	数量	百分比	数量	百分比	数量	百分比
本文方法	44	97.8%	42	93.3%	43	95.6%
基于单独关键帧全局颜色特征的方法	25	55.6%	28	62.2%	26	57.8%

上述实验结果显示,基于“场景词汇”的故事单元相似度计算方法与用户的主观判断具有较好的一致性。通过用户反馈信息的比较,“场景词汇”作为一种中间层次的特征,相比传统的颜色特征等底层特征而言,是一种用户能够比较准确地感受和描述的语义特征;此外,基于“场景词汇”的故事单元相似度计算方法综合利用了故事单元内部和故事单元之间镜头关键帧的编排信息,是表示一种全局层次的相似程度,而基于镜头关键帧全局颜色特征的故事单元相似度计算方法,则是选择出故事单元之间各对关键帧相似度的计算结果的最大值来表示故事单元的相似程度,信息比较片面,因此,基于“场景词汇”的故事单元相似度计算方法更加能够贴近用户对于相似程度判断的主观感官,可以作为一种有效的信息在各种应用需求中加以充分研究和利用。

4 总结和展望

当前研究中的 NDK 识别信息为多语种和多来源的新闻视频中故事单元的相似度关联提供了良好的基础,针对目前研究缺乏故事单元基于 NDK 识别信息的相似度定量计算和分析的问题,本文研究提出了“场景词汇”的概念,分析了“场景词汇”的特殊属性,设计了一种故事单元基于“场景词汇”的描述和相似度计算方法,该方法以向量空间模型为基础,对权重计算方法等进行了改进。通过实验的数据和反

馈显示, 故事单元中的“场景词汇”能够比较好地贴近用户感官和实际需求。

本文研究的“场景词汇”作为一种比较宏观的视觉特征, 在进一步的研究中可以结合视频标注、图像标注等技术进一步完善, 通过与各种语义层次信息的融合, 能够更好地满足各种应用需求。

参 考 文 献:

[1] Ide I, Mo H, Katayama N, et al. Threading News Video Topics[C]// ACM Workshop on Multimedia Information Retrieval (MIR2003) USA, 2003: 239- 246.

[2] Ide I, Mo H, Katayama N, et al. Exploiting Topic Thread Structures in a News Video Archive for the Semi-automatic Generation of Video Summaries [C]// 2006 IEEE Intl. Conf. on Multimedia and Expo (ICME2006), Canada, 2006: 1473- 1476.

[3] Allan J, Wade C, Bolivar A. Retrieval and Novelty Detection at the Sentence Level[C]// Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2003) Cananda, 2003: 314- 321.

[4] Wu X, Ngo C W, Li Q. Threading and Autodocumenting News Videos: A Promising Solution to Rapidly Browse News Topics[J]. IEEE Signal Processing Magazine, 2006, 23(2): 59- 68.

[5] Peng Y X, Ngo C W, Dong Q J, et al. An Approach for Video Retrieval by Video Clip[J]. Journal of Software, 2003, 14(8): 1409- 1417.

[6] Lin T, Zhang H J, Feng J F, et al. Shot Content Analysis for Video Retrieval Applications[J]. Journal of Software, 2002, 13(8): 1577- 1585.

[7] Zhang D Q, Chang S F. Detecting Image Near-duplicate by Stochastic Attributed Relational Graph Matching with Learning[C]// Proceedings of ACM Multimedia, 2004: 877- 884.

[8] Chang S F, Hsu W, Kennedy L, et al. Columbia University Trecvid-2005 Video Search and High-level Feature Extraction[C]// Proceedings of Trecvid Workshop, 2005.

[9] Chua T S, Neo S Y, Zheng Y T, et al. Trecvid-2006 by NUS-I2R[C]// Proceedings of Trecvil Workshop, 2006.

[10] Hsu W, Chang S F. Topic Tracking across Broadcast News Videos with Visual Duplicates and Semantic Concepts[C]// Proceedings of International Conference on Image Processing, 2006, USA.

[11] Zhai Y, Shah M. Tracking News Stories across Different Sources[C]// Proceedings of the 13th Annual ACM International Conference on Multimedia (ACM MM2005) Singapore, 2005: 2- 10.

[12] 文军, 吴玲达, 曾璞, 等. 多源新闻视频中相似关键帧分析研究[J]. 小型微型计算机系统, 2009, 30(4): 770- 774.

[13] Cheung S C, Zakhor A. Efficient Video Similarity Measurement with Video Signature[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(1): 59- 74.

[14] Jain A K, Vailaya A, Xiong W. Query by Video Clip[J]. ACM Multimedia Systems, 1999, 7(5): 369- 384.

[15] Odobez J M, Perez D G, Guillemot M. Video Shot Clustering Using Spectral Methods[C]// Third International Workshop on Content-based Multimedia Indexing (CBMI 2003) France, 2003: 94- 102.

[16] Grauman K, Darrell T. Efficient Image Matching with Distributions of Local Invariant Features[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2005 (CVPR 2005), USA, 2005: 627- 634.

(上接第 114 页)

参 考 文 献:

[1] Torrieri D J. Statistical Theory of Passive Location Systems[J]. IEEE Transactions on Aerospace and Electronic Systems, 1984, 20(2): 183- 198.

[2] 孙仲康, 周一宇, 何黎星. 单多基地有源无源定位技术[M]. 北京: 国防工业出版社, 1996.

[3] Zhou Y Y. Analysis of Location Accuracy for an Emitter Using Satellite-mounted Interferometer[J]. Chinese Journal of Aeronautics, 1998(1): 29- 36.

[4] 魏星, 万建伟, 皇甫堪. 基于长短基线干涉仪的无源定位系统研究[J]. 现代雷达, 2007, 29(5): 22- 35.

[5] Gavish M, Weiss A J. Performance Analysis of Bearing-only Target Location Algorithms[J]. IEEE Transactions on Aerospace and Electronic Systems, 1992, 28(3): 817- 828.

[6] Levanon N. Lowest GDOP in 2-D Scenarios[J]. IEE Proc. Radar Sinar Navig., 2000, 147(3): 149- 155.

[7] 龚文斌, 谢恺, 冯道旺, 等. 星载无源定位系统测向定位方法及精度分析[J]. 长沙电力学院学报(自然科学版), 2004, 19(2): 64- 71.

[8] Spingarn K. Passive Position Location Estimation Using the Extended Kalman Filter[J]. IEEE Transactions on Aerospace and Electronic Systems, 1987, 23(4): 558- 567.

[9] 郗晓宁, 王威, 高玉东. 近地航天器轨道基础[M]. 长沙: 国防科技大学出版社, 2003.