ISSN 1000-0054 清华大学学报（自然科学版）2007 年 第 47 卷 第 1 期 29/41
CN 11-2223/N J Tsinghua Univ (Sci & Tech), 2007, Vol.47, No.1 112-115

基于 HHMM 的多线索融合和事件推理方法

金国英, 陶霖密, 徐光祐, 张 翔

(清华大学 计算机科学与技术系, 北京 100084)

摘 要: 为了解决基于内容检索技术中低层特征与高层语义之间存在语义间隔问题提出了基于多层次线索与事件的分层模型, 以及相应的基于分层隐Markov 模型(HHMM)的多线索融合和事件推理方法。其中线索是对事件进行推理的要素, 它是低层特征与事件之间的中间层次。在将视频流分割为镜头后, 从各个镜头中抽取若干与事件密切相关的线索, 构造并训练各事件的 HMM 模型, 用于融合线索和进行事件推理。由于输入视频通常包含多个事件, 不可避免会遇到时域分割问题, 因此构造一个 HHMM 模型用于同时进行视频流的合理分割和事件的识别。对足球视频的大量实验结果表明, 该方法可有效地检测足球视频事件, 并在抽取的线索不完全可靠的情况下具有一定的鲁棒性。

关键词: 模式识别; 视频处理和分析; 基于内容检索; 事件检测; 分层隐Markov 模型(HHMM)

中图分类号: TP 391.4 **文献标识码**: A

文章编号: 1000-0054(2007)01-0112-04

Cue fusion and event inference based on HHMM

JIN Guoying, TAO Linmi, XU Guangyou, ZHANG Xiang

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

Abstract: A cues fusion and events inference method was developed based on the hierarchical hidden Markov model (HHMM) to bridge the semantic gap between the low-level features and the high-level semantics in content-based retrievals. Cues are introduced into the system as an element for inferring higher-level events. In the system framework, the input video stream is first segmented into shots, then, semantic cues are extracted from the shots based on low-level features, and HHMM models are built and trained to infer the events from the cues. The input video stream usually contain more than one event, so a temporal segmenting video stream is used to segment events for the HHMM-based events inference. An HHMM model was developed to group shots and to recognize simultaneously events in a soccer video. Tests on the soccer videos show that the system is effective and robust in inferring events from roughly extracted cues.

Key words: pattern recognition; video processing and analysis; content based retrieval; event detection; hierarchical hidden Markov model

基于事件的分析方法是目前基于内容检索技术的研究热点, 其基本思想是根据上下文和领域知识从低层特征中抽取线索, 并融合各线索推理出高层的语义事件。已有不少以体育比赛为实验数据的高层语义研究[1,2], 其中事件推理方法包括利用一些启发式规则[2], 或引入诸如动态Bayesian 网(DBN)或隐Markov 模型(HMM)等概率模型。

最常见的基于 HMM 推理的事件检测方法一般为每个事件构造一个 HMM 模型, 用最大似然分类器找出输入视频所属的事件。但这种方法不能处理输入视频中包含多个事件的情况。如何将视频分割成合适的片段, 使得每个片段产生的观察值序列必定符合某个 HMM 事件模型, 称之为基于 HMM 的事件检测方法中的时域分割问题, 即视频分割和事件识别的冲突问题。为解决这一问题, 目前采用的方案可分为先分割后识别、分割与识别同时进行这两大类。前者[3]的分割结果好坏很大程度上会影响事件识别结果。后者中较常用的方法是引入多层次 HMM 模型[4]。

本文引入分层HMM (hierarchical HMM, HHMM)模型来解决时域分割问题。与目前常见的事件检测方法不同的是, 用线索来代替底层特征进行 HMM 推理, 一方面降低了 HMM 的状态数量, 简化了 HMM 的训练和推理过程。另一方面, 由于一些线索具有一定的语义, 使得 HMM 推理过程与直接用底层特征进行推理相比更加合理。线索的引入为高层语义的提取提供了一个多层次的通用模型。从特征得到的最低层的线索, 可以推理出语义层次较低的事件, 这些子事件又可以作为线索来推理高层

收稿日期:2005-11-25

基金项目: 国家自然科学基金资助项目(60673189);
中国博士后科学基金资助项目(2005038351)

作者简介: 金国英(1974-), 女(汉), 上海, 博士研究生。

通讯联系人: 陶霖密, 副教授, E-mail: linmi@tsinghua.edu.cn

次事件, 以此类推, 构成一个基于多线索和事件的分层模型。在各层次用线索推理出事件时可采用不同的方法, 如利用启发式规则、DBN 和 HMM 等。

1 足球视频事件的 HMM 模型

1.1 足球比赛视频的语法分析

足球比赛视频是按照特定的风格生成的, 称之为足球比赛视频的语法。比赛的一般过程是由显示赛场全景的远镜头和包含球员动作的中镜头互相交错而构成的, 中间可能夹杂几个对球员、教练或裁判的特写镜头(近镜头)。在特殊事件发生后才会出现慢镜头回放, 因此, 在选择待检测事件时首先考虑这些被回放的事件: 射门事件和犯规事件。

射门事件的第一个镜头是包含射门动作的远镜头, 然后是一个或多个从不同角度对射门场景的回放, 这些回放镜头大多数是中镜头。射门球员或守门员的特写镜头有可能出现在回放镜头的前后。直至出现一个新的远镜头或开球的中镜头, 一个完整的射门事件结束。

类似地, 犯规事件从包含犯规动作的远镜头或中镜头开始, 接着是犯规动作的中镜头回放, 回放前后可能有犯规球员的特写、裁判进行判罚的场面等, 直至出现一个新的远镜头或开球的中镜头。

1.2 足球视频事件的 HMM 模型

根据上文中对足球视频的语法分析可以看到, 在射门和犯规事件的发生过程中, 镜头内容随时间的变化具有一定规律, 这样的规律可以用 HMM 模型进行建模, 不同的镜头内容对应于 HMM 模型中的各个状态。

通过对足球视频事件的分析, 定义了射门和犯规事件的 HMM 模型结构(见图 1)。其中, 状态 GV 代表远镜头, ZI 代表中镜头, CU 代表特写镜头, RP 代表回放镜头。射门和犯规事件就是由这些镜头按照各自的规律组合起来的。在这两个事件中, 镜头内容的变化规律有很强的时序性, 每个状态只可能由处于其左侧的状态转移过来。因此, 选择左-右 HMM 模型来为射门和犯规事件建模。

此外, 还定义了一般比赛进程事件来表示除了射门和犯规这两个特殊事件之外的其他事件。一般比赛进程中没有回放镜头, 而且镜头内容的变化没有什么时序关系, 因此采用包含 GV、ZI 和 CU 共 3 个状态的全连接 HMM 模型。

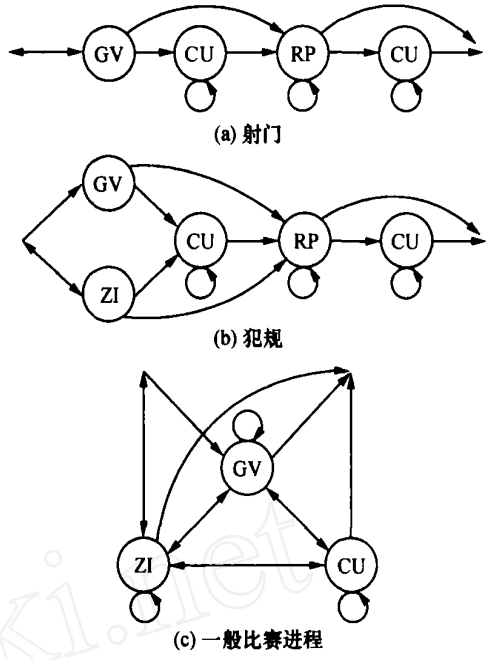

图 1 3 个事件的 HMM 模型

2 线索的选择和生成

线索是连接特征和事件的中间层次, 由底层特征计算或推理而成, 经过推理产生更高层次的语义信息(如事件)。本文中镜头是线索生成和事件检测的基本单位, 因此需要先进行镜头分割[5], 再根据各个镜头中的底层特征计算或推理出线索。

在选择线索时需要权衡线索与事件的相关性和提取的可行性这两个方面。一方面, 根据事件的推理过程, 选择与事件紧密关联的线索。另一方面, 需要验证这些线索能否用现有的技术自动提取。根据这两个准则, 选择了 7 个线索用于事件推理。

1) 镜头尺度: 远/中/近。整个图像域的草地颜色比[1]是判别镜头尺度的重要特征之一。

2) 慢镜头: 慢镜头可能由一般摄像机拍摄, 插入重复帧或插值帧后播放; 也可能由高速摄像机拍摄, 然后按常速播放。基于一种较为通用的基于双层 HMM 的慢镜头分割方法[5], 并结合了镜头分割结果, 将慢镜头检测问题转化为镜头转换片断的分割和识别问题。

3) 球门和球网: 在中/近镜头中利用球网的分形纹理特征; 在远镜头中, 则采用基于 Hough 变换的门柱和场地平行线检测方法。

4) 标题条: 标题条的出现与特殊事件的发生有一定的关联。综合利用了标题条的时域和空域特性, 例如稳定性、球场颜色特性、字符的强水平边缘特性等等来检测和定位标题条。

5) 人体区域: 对人体区域的检测综合利用了人脸和服装颜色信息。

6) 镜头模糊度: 镜头模糊度在一定程度上体现了图像中运动的剧烈程度。使用Marr-Hildreth过零算法对整幅图像进行边缘检测,边缘点的数量可以用来估计模糊度。

7) 镜头频度: 在一定程度上反映了比赛的精彩程度。一种与镜头频度成反比的度量是镜头平均帧数,即计算以此镜头为中心的给定长度窗口内镜头帧数的平均值。

3 基于 HHMM 的视频事件检测方法

基于HMM 的事件检测方法中, 输入视频所属的事件是由最大似然估计(MLE)分类器确定的最适合于观察值序列的HMM 模型所决定的。但是这种方法不能处理输入视频中包含多个事件的情况。为此,引入HHMM 模型来处理HMM 应用所面临的时域分割问题。

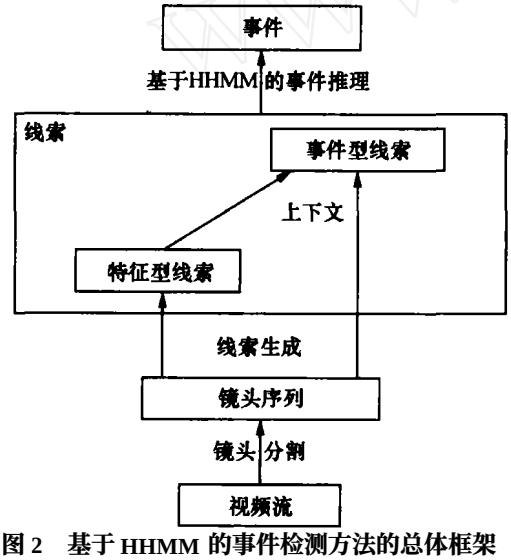

图2 基于 HHMM 的事件检测方法的总体框架

图2为基于HHMM 的事件检测方法的总体框架。本文中线索生成和事件推理都是以镜头为基本单位,因此,需要首先将视频流分割为镜头序列。从每个镜头中生成各线索(包括来自于特征的特征型线索和来自于子事件的事件型线索),融合各线索构成HHMM 的观察值向量; 根据产生的观察值序列进行基于HHMM 的事件推理,得到事件检测结果。

3.1 HHMM 建模和训练

HHMM 模型是HMM 模型的一种扩展[6], 主要特点包括:1) 每个高层HMM 状态对应于一个低层的HMM 模型(即子HMM 模型);2) 高层HMM 模型的状态转移只能在低层的子HMM 模型进入终结状态后被激活;3) 观察值只能由最低层的HMM 模型的状态产生。

构造的足球视频的HHMM 模型(如图3所示),高层的HMM 模型是一个3个状态的全连接HMM 模型,其中的3个状态ST、FL 和NP 分别对应于射门事件、犯规事件和一般比赛进程的HMM 模型。HHMM 模型的观察值向量也就是底层HMM 模型的观察值向量,是由第2节中抽取的线索所组成。其中,第1个线索镜头尺度可能的取值有3个: 远/中/近。第2到第5个线索都只有两个可能的取值: 是/否(或有/无)。最后两个线索是连续变量,为了与其他线索保持一致,将这两个线索分别进行归一化并量化为2级,分别对应于取值: 高/低。这样,从图像本身所具有的特征直接计算而得的线索也具有了一定程度的语义。由这些线索组成的HHMM 的七维观察值向量,其可能的取值数量$M=3\times2^6=192$。

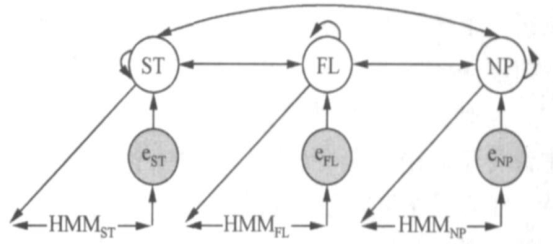

图3 足球视频的 HHMM 模型简图

用有监督的学习方法来训练HHMM 模型参数,即样本数据是经过手工分割和标注的。每个事件的底层HMM 模型都用分割后相应的事件子序列分别进行训练。事件e的底层HMM 模型HMM_e(e= ST, FL, NP)的参数可用Baum-Welch 算法估计出。高层的HMM 状态之间的状态转移概率和层间的状态转移概率可根据标注内容计算出来: 高层HMM 状态之间的转移概率a_{xy}等于样本视频中事件x之后是事件y的数量除以样本视频中事件x的数量; 高层HMM 状态到它对应的HMM 模型HMM_e中各个状态的转移概率,就是HMM_e的初始状态概率; 事件e的各个状态到其终结状态的转移概率a_{ie}, 等于结束状态为θ_i的事件e个数除以事件e的总数。

3.2 HHMM 推理和足球视频事件检测

估计出HHMM 模型参数后,可以用广义Viterbi 算法[6]求出输入视频的最优状态序列,并由此确定输入视频中射门和犯规事件的位置和边界。

一个完整的足球视频事件检测过程按照下列步

骤进行:1) 将输入视频分割成镜头序列;2) 从每个镜头中抽取 7 个线索组成观察值, 则可得到输入视频的观察值序列;3) 将观察值序列送到 HHMM 模型中, 用广义 Viterbi 算法推理出最优状态激活序列;4) 从状态 ST 开始到状态 e_{ST} 结束的视频片段是射门事件, 从状态 FL 开始到状态 e_{FL} 结束的视频片段是射门事件。

4 实验结果

采集了 67 个足球比赛视频, 持续时间共长达 17 h 27 min。这些视频按照来源不同可分成 6 类, 分别来自两届世界杯(WC), 欧洲冠军联赛(UEFA), 意大利甲级联赛(IFLSA), 以及由两个电视台分别转播的英超联赛(EPL)。

各线索作为事件推理的依据, 其检测效果显然会影响事件检测结果。实验表明, 用完全正确的慢镜头线索来进行 HHMM 推理, 可以得到很好的事件检测结果。慢镜头对事件检测率的影响较大, 是由于待检测事件从定义到检测和识别, 都与慢镜头密切相关。其他线索对事件检测率的影响相对小些, 经校验后最多可提升 4% 到 5%。

关注的另一个问题是解决 HMM 应用中遇到的时域分割问题。为此, 提出了 3 种方案进行实验(都用未经校验的线索进行推理), 比较了在不同的时域分割方案下的事件推理的精确度(见图 4)。

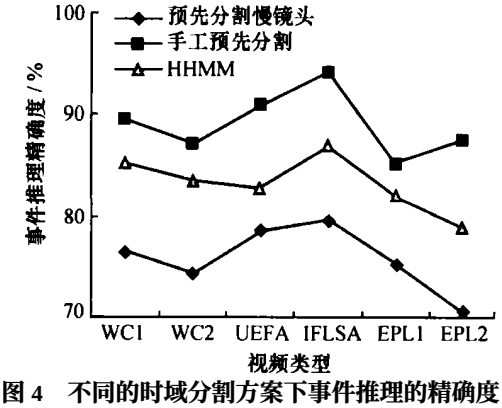

图 4 不同的时域分割方案下事件推理的精确度

手工预先分割的方法[6] 将输入的镜头序列分割成若干镜头组。每个镜头组都对应于某个事件, 即用 ML 分类器找出的那个 HMM 模型对应的事件。由于是手工预先分割, 不会出现事件边界的偏差。此方法事件推理精确度比较高。

基于慢镜头的预先分割方法是在慢镜头检测的基础上, 根据一些判别规则自动地将镜头序列分割成若干镜头组, 后续步骤与手工预先分割方法一致。这种预先分割方法由于非常依赖慢镜头检测的可靠性, 而未经校验的慢镜头线索本身正确率不是很高, 被误分的镜头组产生的观察值序列可能不符合任何一个事件模型, 或者被检测为错误的事件。所以, 事件推理效果较差。

与预先分割方法相比, 基于 HHMM 的方法同时进行分割和识别, 虽然事件推理的精确度不能达到手工分割的高度, 但同样作为自动检测的方法, 比用基于慢镜头的预先分割方法得到的事件推理结果要好得多。因此, 基于 HHMM 的事件推理方法是一个精确度基本上可以接受, 同时又不需要过多的人工干预的事件检测自动方法。

5 结 论

本文提出了基于 HHMM 的多线索融合和事件推理方法, 并实际应用于体育视频分析。本方法引入的多线索这个中间层次, 将原有的难以跨越的从底层特征到高层语义之间的鸿沟, 转化为从底层特征检测线索和从线索推理高层语义这两个可以解决的问题。同时, HHMM 的引入解决了 HMM 应用中的时域分割问题。对大量足球视频的实验证明, 基于 HHMM 的多线索融合和事件推理方法可以将射门和犯规等事件自动地检测出来。

参考文献 (References)

[1] Xu P, Xie L, Chang S F, et al. Algorithms and systems for segmentation and structure analysis in soccer video [C]// Proceedings of IEEE International Conference on Multimedia and Expo. Tokyo, Japan: IEEE Press, 2001: 928-931.

[2] Tovinkere V, Qian R J. Detecting semantic events in soccer games: towards a complete solution [C]// Proceedings of IEEE International Conference on Multimedia and Expo. Tokyo, Japan: IEEE Press, 2001: 833-836.

[3] Xie L, Chang S F, Divakaran A, et al. Structure analysis of soccer video with hidden Markov models [C]// Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. Orlando: IEEE Press, 2002: Ⅳ 4096-Ⅳ 4099.

[4] Xu G, Ma Y F, Zhang H J, et al. An HMM based semantic analysis framework for sports game event detection [C]// Proceedings of IEEE International Conference on Image Processing. Barcelona: IEEE Press, 2003: Ⅰ 25-Ⅰ 28.

[5] Jin G Y, Tao L M, Xu G Y. S low-motion replay detection in soccer videos based on multi-level HMM integrated with shot detection [C]// Proceedings of Workshop on Image Analysis for Multimedia Interactive Services. Montreux, Switzerland: IEEE Press, 2005.

[6] Fine S, Singer Y, Tishby N. The hierarchical hidden Markov model: Analysis and applications [J]. *Machine Learning*, 1998, **32**(1): 41-62.