

一种足球视频精彩事件检测方法

毕殿杰, 陈 涛

(安徽财经大学 信息工程学院, 安徽 蚌埠 233041)

摘 要:限于当前的技术水平, 视频检索技术难以在底层特征与高层语义之间建立通用的视频分析模型。文中结合足球视频的领域知识, 着重分析了一类特殊的语义事件——精彩事件, 基于统计的方法提出了动态贝叶斯网络事件检测模型, 以及相应的学习和推理算法。实验结果表明, 该方法可有效地提取足球视频中的精彩语义事件, 具有较高的查全率和查准率, 较强的鲁棒性, 是一种很有前景的视频语义事件检测方法; 同时证明了, 通过结合某一领域知识, 底层特征与高层语义之间是可以建立起某种联系的。

关键词:视频检索; 精彩事件; 动态贝叶斯网络; 领域知识

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2010)05-0219-04

A Soccer Video Highlights Detection Method

BI Dian-jie, CHEN Tao

(School of Information Engineering, Anhui University of Finance & Economics, Bengbu 233041, China)

Abstract: Limited to the current technical level, it is difficult for sport video analysis technology to establish universal video analysis model between the bottom characteristics and the high-level semantics. Through a combination of the corresponding field of knowledge, it is likely to establish links between the bottom characteristics and the high-level semantic. Combine domain knowledge of soccer video, focusing on an analysis of a special kind of semantic events, i.e. highlights. Based on statistical methods proposed dynamic Bayesian network model of highlights detection, and the corresponding events of learning and reasoning. The experimental results show that the method can effectively extract soccer video highlights. Has a high recall rate and precision rate, strong robustness, it is a promising video semantic event detection method. At the same time proved that through a combination of a particular field of knowledge, the underlying characteristics and high-level semantics is possible to establish a link.

Key words: video retrieval; highlight; dynamic Bayesian network; domain knowledge

0 引言

对视频内容分析与检索技术的研究涉及到人工智能、计算机视觉、信号处理、模式识别、数据挖掘等诸多学科领域, 具有重要的理论和现实意义, 其最终目标就是帮助人们从海量数据中方便、快捷地找到需要的视频资源。巨大的商业前景和重要的学术价值, 吸引了众多的研究人员在这一问题上开展研究。一些系统先后被提出, 这些系统主要是通过提取和索引视频中的颜色、纹理、形状和运动等底层感知特征来实现视频检索, 还不能提供一般用户所需要的高层语义查询功能^[1]。限于当前的技术水平, 难以建立底层特征与高

层语义之间通用的视频分析模型。但是通过结合相应的领域知识, 是可能将底层特征与高层语义建立某种联系的。

基于此, 结合足球的领域知识, 提出了一种基于动态贝叶斯网络模型的精彩事件提取方法。这里的精彩事件表示比赛中观众比较感兴趣的片断, 比如射门、角球、任意球、犯规事件, 以及运动员精彩的过人动作等。在早先的相关工作中, 已经有一些文献开始探讨基于统计的体育视频分析方法, 较常用的是贝叶斯网络 (Bayesian Network, BN) 和隐马尔科夫模型 (Hidden Markov Model, HMM)。这些方法在应用到视频分析的时候, 都存在一定的局限性。贝叶斯网络能够很好地进行分类, 但它是一个静态分类模型不能充分利用随时间变化的上下文信息。隐马尔科夫模型适于对时间信号的处理, 但是在视频内容分析中其模型表达能力还是有限的, 这主要是因为视频是一种既有空间信息又有时间信息的多维信号。文中引入了一类更为强

收稿日期: 2009-08-23; 修回日期: 2009-11-05

基金项目: 安徽省自然科学基金(KJ2009B060Z); 安徽财经大学09年度青年科研项目(ACKYQ0947ZC)

作者简介: 毕殿杰(1977-), 男, 讲师, 研究方向为人工智能、多媒体技术等; 陈 涛, 副教授, 研究方向为数据挖掘等。

大的时序信号统计工具——动态贝叶斯网络(Dynamic Bayesian Network, DBN),用于足球视频的内容分析。一方面,通过考虑各个时刻间的转移概率,动态贝叶斯网络扩展了贝叶斯网络对时序信号的建模能力。另一方面,动态贝叶斯网络允许在同一个时间点上使用多个状态变量,而不是像隐马尔科夫模型那样只有一个状态变量^[2]。

1 动态贝叶斯网络的优势

基于统计的贝叶斯网络,本质上是一个有向无循环的图表模型,直观地表述了多个变量之间的概率依赖关系;并且,贝叶斯网络可以有效地把先验知识和现有数据结合起来,使得网络的推理结果更加的合理。贝叶斯网络有两个基本要点:学习和推理。学习,是指根据系统部分或全部节点的可观察行为来估计系统的概率分布。推理,是指已知系统的概率分布,根据部分节点的状态推导其他节点的状态。

动态贝叶斯网络是贝叶斯网络在时间变化过程上的扩展,反映了一系列变量随时间变化的情况。为方便处理,假设动态贝叶斯网络满足两个条件:1)网络拓扑结构不随时间发生变化,即除去初始时刻,其余时刻的变量及其概率依存关系完全相同;2)满足一阶马尔可夫条件,即给定当前时刻变量的状态后,未来时刻的状态和先前时刻的状态无关。满足上述条件后,动态贝叶斯网络可以看做是贝叶斯网络在时间序列上的展开^[3],如图 1 所示。

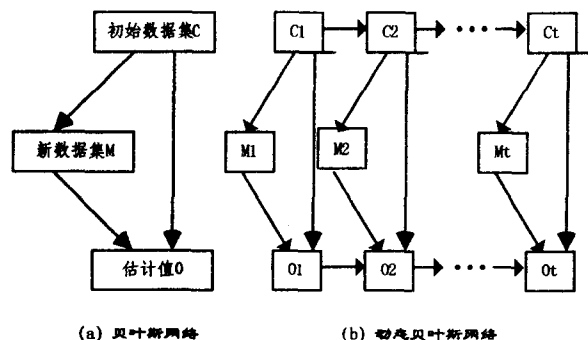


图 1 网络结构

考虑图 1 (b) 所示的动态贝叶斯网络,其描述了变量集 $X_t = \{C_t, M_t, O_t\}$ 的概率依存关系及其随时间 $t = 1, \dots, T$ 变化的情况。在任意时刻 t , 变量 M_t 的状态由变量 C_t 决定,而 O_t 的状态则由 C_t 和 M_t 共同决定,即变量集 X_t 的联合概率分布为:

$$P(X_t) = P(C_t, M_t, O_t) = P(C_t)P(M_t | C_t)P(O_t | C_t, M_t)$$

考虑 O_t 与 C_t 之间的条件概率分布,在时刻 $t-1$ 和 t 之间,变量 C_t 的状态发生了转移,因此变量集 X_t

的转移概率为:

$$P(X_t | X_{t-1}) = P(C_t | C_{t-1})$$

可以看出,动态贝叶斯网络通过网络拓扑结构反映变量间的概率依存关系及其随时间变化的情况,其不但能够对变量所对应的不同特征之间的依存关系进行概率建模,而且对特征之间的时序关系也能很好地加以反映,因此适合于对视频这种同时具有特征相关性和时序相关性的复杂特征进行联合建模。

除此之外,动态贝叶斯网络还可任意改变拓扑结构或增删变量以反映变量间各种不同的关联关系,而不影响训练算法本身,因此具有很好的可扩展性和灵活性。动态贝叶斯网络还具有良好的可解释性,其拓扑结构具有精确及易于理解的概率语义,通过对其进行分析可以加深对不同变量间关联关系的理解。

在实际应用中,使用动态贝叶斯网络便于结合先验信息(领域知识)对模型进行设计。通常,动态贝叶斯网络的节点可以直接从现实系统中对应过来。节点之间的因果关系,即条件概率,可以由专家设定,也可以从训练数据中学习得到。此外,在动态贝叶斯网络中,节点之间并不需要完全连接。事实上,有些节点之间是没有直接联系的,即是条件独立的;而条件独立性有助于降低联合概率的计算量,这恰好是动态贝叶斯网络的一大优势。

2 领域建模

在足球视频中,重点关注五种精彩事件:射门、角球、任意球、过人动作、犯规,这些精彩事件都是具有丰富的高层语义。很显然,从纹理、形状、颜色等底层特征直接映像到这些高层语义事件是非常困难的。但考虑到,精彩语义事件是由各个场景组成的,这些场景可映射到底层特征上。在这种间接映像方式中,底层特征与场景之间能够形成比较有效的映射,再通过对场景的统计推理,就可以检测和识别出语义事件^[4,5]。

基于这种思想,根据足球比赛的规则和电视转播的规范,采用了一种场景分类方法。该方法对于每一帧足球视频,依赖于球场检测分割的结果,提取了球场比率、水平和垂直投影直方图,以及球场边界形状特征;将足球场景分成场内和场外两种,进而将场内场景分为局部场景和全局场景。全局场景又可以分为,球门区域、角球区域、中场区域等。这种分类可以给足球视频的高层分析提供十分明确的语义信息,例如,球门区域场景的出现很可能与“射门”等精彩事件有关。既然这些场景画面和所要分析的精彩语义事件之间存在紧密的联系,则根据这些场景之间的上下文关系和转移概率就可以判断事件的发生。笔者还对动态场景进

行了分类。主要是基于估计的摄像机参数,将摄像机运动分为“静止”、“快速水平平移”、“快速竖直平移”、“快速缩放”等。因为在足球视频中,摄像机大多数时间集中在运动员或者足球上,对于摄像机运动的分类同样可以为足球视频的高层内容分析提供十分明确的语义信息。通过该场景分类方法,可以把足球视频的镜头这一低层结构单元提升到场景这一中间层结构单元上,继而为足球视频的高层内容分析所服务^[6]。

基于以上分析,构造了足球视频的分层动态贝叶斯网络模型,如图2所示。该模型自上而下依次是事件层、场景层和观测层。通过这种三层网络,底层特征以一种统计推理的方式映像到高层语义事件上。

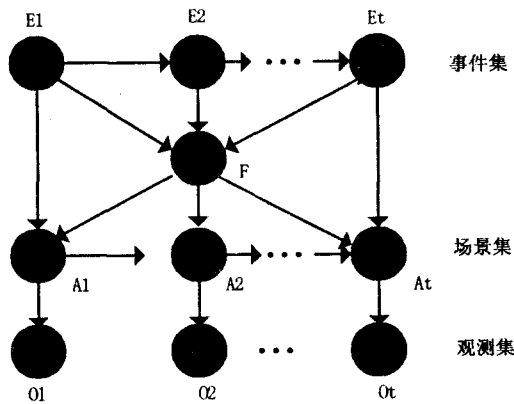


图2 足球视频内容分析模型

模型高层是一个5个状态的全连接模型,其中的5个状态分别对应于射门事件、犯规事件、角球事件、任意球事件和过人动作,由随机变量 $E_t = \omega_i, i = 1, \dots, 5$ 表示。这里假定环境满足马尔可夫模型,也就是说,在 $t+1$ 时刻,环境状态所能知道的东西(没有感知的)完全取决于 t 时刻的环境状态和动作(概率地),条件独立于在 t 时刻以前的每个事情。模型中间层是基本场景 A_t 层,而每个待检测的精彩事件都是由这些基本场景组成的。模型底层是观测值向量,观察测向量是由抽取的特征线索所组成的,也就是说从视频帧中提取特征作为模型观测输入以区分不同的场景。文中特征线索生成和事件推理都是以镜头为基本单位,因此,首先需要将视频流分割为镜头序列;其次从镜头序列中,对每帧图像的特征进行提取,生成特征线索构成观测值向量;根据产生的观测值序列进行事件推理,得到事件检测结果^[7]。

有时,网络结构可以通过增加隐藏节点而大大地简化,该模型中的节点 F 就属于隐藏节点,引入它既可以大大减少计算量,又可以起到事件指示开关作用,引导事件转移。当随着时间的推进属于事件 E_t 的所有场景已经完成时,开关打开表示可以转移到下一个事

件。

3 模型学习和推理

3.1 学习

模型学习的任务是根据给定的训练数据来估计模型的参数。

在观测层,给定用于训练的观测序列 $O_{1:T}$,模型的参数 Θ 可以由最大似然方法估计,即

$$\hat{\Theta} = \arg \max_{\Theta} P(O_{1:T} | \Theta)$$

Expectation-maximization (EM) 算法具有线性的收敛速度,并且简单,容易实现,可以使似然函数达到局部最优。故文中采用 EM 算法来进行迭代求解。

$$\Theta^{(n+1)} = \arg \max_{\Theta} E\{\log P(O_{1:T}, A_{1:T} | \Theta) | O_{1:T}, \Theta^{(n)}\}$$

其中 $\Theta^{(n)}$ 是第 n 次的迭代结果。通过迭代计算,参数可收敛到一个局部极值点。

在模型中间情景层,节点间的条件概率分布可以从事件 ω 的训练样本中直接统计得到。设训练样本中事件 ω 的数目为 N ,其中起始基元为 σ 的有 M 个,则事件 ω 中起始基元为 σ 的概率:

$$P(A_t = \sigma | A_{t-1} = \sigma', E_t = \omega, F = 1) = \pi_{\omega}(\sigma) = \frac{M}{N}$$

设所有事件 ω 的样本中基元 σ' 的数目为 S 个,其中后面紧接 σ 的为 T 个,则事件 ω 中基元 σ' 到 σ 的转移概率:

$$P(A_t = \sigma | A_{t-1} = \sigma', E_t = \omega, F = 0) = A_{\omega}(\sigma', \sigma) = \frac{T}{S}$$

在模型事件层,事件节点相关的条件概率分布也可以从样本数据中各类事件发生的统计中得到。

$$P(E_1 = \omega) = \pi(\omega)$$

$$P(E_t = \omega | E_{t-1}) = A(\omega', \omega)$$

式中的 π 表示序列起始事件的概率分布, A 表示事件之间的转移概率分布^[8]。

3.2 推理

在估计出三层动态贝叶斯网络模型参数后,就可以进行推理,推理过程实际上是一个精彩事件的识别过程。给定 C 个已经训练好的模型 M_1, M_2, \dots, M_C ,其中每个模型对应一类精彩事件^[9]。给定一个测试行为,其观测序列为 $O_{1:T}$,则选择 C 作为该测试行为的类别标号,其中 C 由下式确定:

$$C = \arg \max_i P(M_i | O_{1:T}) = \arg \max_i P(O_{1:T} | M_i) P(M_i)$$

式中 $P(M_i)$ 是模型 M_i 的先验概率。假设模型 M_i

的参数是 Θ_i , $P(O_{1:T} | M_i) = P(O_{1:T} | \Theta_i)$ 。

4 实验与分析

算法评估采用以下的评价指标:

1)查全率:即精彩事件的检出率,反映该模型算法发现精彩事件的能力。当查全率越大时,漏掉的精彩时间数越少,算法性能也越好。

2)准确率:即精彩事件的检对率,反映“找对”精彩事件的能力。准确率越高,系统性能越好。

在实验中,考虑到用户对自动提取精彩片断的需求,把视频中的射门、角球、任意球、精彩过人、犯规都看作是精彩片断,以评估算法提取精彩片断的性能。测试和训练数据来自于 08-09 赛季欧洲冠军杯的五场比赛的视频,总长度约 470 分钟。使用其中的一半作为训练集,另一半作为测试集;并事先对视频片断中的精彩事件均进行了人工标注作为真实参考数据。

表 1 精彩事件检测的实验结果

精彩事件	正确	错误	丢失	查全率	查准率
射门	49	12	9	84.48%	80.33%
角球	28	12	4	87.50%	70.00%
任意球	15	6	5	75.00%	71.43%
过人动作	19	12	11	63.33%	61.29%
犯规	21	10	7	75.00%	76.74%
总计	132	52	36	78.57%	71.74%

表 1 给出了根据动态贝叶斯模型检测精彩事件片断的实验结果。算法对“过人动作”这一精彩事件的检测效果较差,事实上,以人工的方式也很难确定一个精彩过人动作;对“角球”和“角球”的检测效果比较理想;对“任意球”和“犯规”的检测效果一般。算法对文中重点讨论的五种精彩事件,平均查准率为 71.74%,平均查全率为 78.57%。为了进一步提高性能,需要有定义更精细的场景,以及更加有效的特征提取和场景识

别算法。

5 结束语

文中结合足球视频领域知识,提出并实现了一种基于动态贝叶斯网络的精彩事件检测模型,及其学习和推理算法。从实验结果可以证明该模型算法是可行的,但是在对个别类型的精彩事件检测上存在一定的误差,在算法设计上需要进一步的改进与完善。

参考文献:

- [1] Flickner M, Sawhney H, Niblack W. Query by Image and Video Content: The QBIC System[J]. IEEE Computer, 1995,28(9):23-32.
- [2] 孙中伟,张福炎.自动视频内容分析综述[J]. 计算机科学, 2002,29(5):80-84.
- [3] 胡静,刘志镜.基于动态贝叶斯网络的多人跟踪算法[J]. 计算机工程,2008,34(16):247-248.
- [4] 薛建儒,郑南宁,钟小品,等.视感知激励——多视觉线索集成的贝叶斯方法与应用[J]. 科学通报,2008,53(2):172-182.
- [5] 杨有,尚晋.一种政府资源档案图像的二值化方法[J]. 计算机科学,2007,34(3):227-229.
- [6] Yu J Q, Wang X, He Y F. Camera calibration and player detection in 3D reconstruction[J]. Journal of Software, 2008, 19(s):151-160.
- [7] 金国英,陶霖密,徐光祐,等.基于 HHMM 的多线索融合和事件推理方法[J]. 清华大学学报:自然科学版, 2007,47(1):112-115.
- [8] 王扉.体育视频的内容分析技术研究[D].北京:中国科学院研究生院,2005:43-58.
- [9] 杜友田,陈峰,徐文立.基于多层动态贝叶斯网络的人的行为多尺度分析及识别方法[J]. 自动化学报,2009,35(3):1-3.
- [8] Harris M, Silvertand J. ArcSDE for Oracle Administration [C]// Twentieth Annual ESRI International User Conference. [s.l.]:[s.n.],2000.
- [9] 傅俊,刘南,刘仁义.实现 ArcSDE 向 Oracle9i Spatial 空间数据的转换[J]. 计算机应用研究,2003(9):99-101.
- [10] 邵立志,杨联安.基于 ArcSDE 和 Oracle 空间数据集成应用研究[J]. 测绘标准化,2008,24(2):22-25.
- [11] 张伟.基于 ArcSDE 和 ArcEngine 的集输管线官网系统的设计与实现[D].西安:西北大学,2008.
- [12] 翁和霞.基于 Oracle9i 和 ArcSDE 的城市基础地理数据库配置研究[J]. 测绘与空间地理信息,2007,30(6):68-70.

(上接第 218 页)

- 2006,23:230-232.
- [3] 何涛,张世禄.基于 ArcGIS 的县级林业资源管理信息系统研究[J],2009,19(2): 183-186.
- [4] 袁明.基于 ArcEngine 的地下管线综合应用系统的设计与实现[D].成都:成都理工大学,2008.
- [5] ESRI 中国(北京)培训中心. ArcGIS Engine 轻松入门教程 [M].北京:ESRI 中国(北京)培训中心,2008.
- [6] ESRI 公司. ArcGIS Developer Help[M]. [s.l.]:ESRI Product Document,2004.
- [7] 刘金龙.基于 ArcGIS Engine 开发的大连市路灯地理信息系统[D].大连:大连理工大学,2007.