

基于 SVM 的视频检索系统框架

赵颖川 王万良 蒋一波 (浙江工业大学 计算机科学与技术学院 浙江 杭州 310013)

摘要: 底层特征到高级语义的转化是视频检索中的重要环节。结合在视频检索中的相关反馈,结果排序以及 BOOST 思想,提出了针对视频检索的改进的 SVM 算法流程,并根据改进的 SVM 算法提出了基于 SVM 的视频检索系统框架。通过采用 TREC2007 的数据测试表明,该视频检索系统框架可达到较高的查准率和查全率。

关键词: MPEG-7; CBVR; 结果排序; 支持向量机 (SVM); 相关反馈

A Video Retrieval System Framework Based on SVM

ZHAO Ying-Chuan, WANG Wan-Liang, JIANG Yi-Bo

(Computer College, Zhejiang University of Technology, Hangzhou 310013, China)

Abstract: Transformation from feature to semantic info is the important step of Video Retrieval. Based on the improved SVM procedure for Video Retrieval according to relevance feedback, reranking and BOOST, this paper proposes a framework. Using the data of TREC2007, the experimental result shows the framework has sufficient improvement in AP and MAP.

Keywords: MPEG-7; content-based video retrieval (CBVR); reranking; support vector machine (SVM); relevance feedback

1 引言

随着视频编解码以及网络的迅速发展,视频文件的数量飞速增长。传统的检索管理方式已无法适应含有大量数据的视频文件,在此背景下,CBVR(Content-based Video Retrieval)成为了一个研究的热门领域。

目前的 CBVR 系统大多基于关键帧进行处理,并进一步根据底层特征通过分类算法进行视频内容的自动标签。常用的算法有朴素贝叶斯, SVM, KNN, MRF(Markov Random Field)等。其中 KNN 当视频数据量较为庞大或者底层特征向量较多的时候计算速度会较慢,朴素贝叶斯算法实现简单,速度比较快,但是对于数据有相关性的时候分类效果较差,而 SVM 不仅可以解决线性不可分的问题,而且需要较小的样本,且不需样本分布的先验知识,同时采用计算速度较快,所以比较适合视频检索系统使用。

支持向量机(Support Vector Machine, SVM)^[1]是一种基于统计学习理论的机器学习方法,它根据 VC 维理论和结构风险最小化原则,通过构造具有最大分类间隔的超平面,最小化泛化误差上限,使得对未知样本的分类具有最优的推广能力。

以往基于 SVM 的视频检索系统都是通过增加 SVM 分类器,如加州大学的 Cuvit^[2]采用了多达 374 个 SVM 分类器,法国信息研究院的 LIG 系统采用了 36 个 SVM 分类器等,或增加 SVM 中使用的底层特征矢量,如来进行系统的优化,而没有将 SVM 的思想融入到视频检索系统设计的过程中,这样的方法不仅会导致查全率的下降,还会大大增加系统的开销。本文结合相关反馈,结果排序以及 BOOST 思想,提出了适合视频检索系统 SVM 算法,利用相关反馈中正反馈和负反馈的特性,以及 BOOST 算法的思想,改进的 SVM 算法引入了权重的概念来对于不同的 SVM 模型

基金项目:浙江省教育厅基金(Y200803365);浙江省新苗人才计划(2008R40G2020003)

收稿时间:2009-10-23;收到修改稿时间:2009-11-28

进行训练,并根据多个 SVM 模型的结果以及权重决定最终的输出。同时利用 SVM 中的超平面方程对于用户的搜索结果进行了排序,通过试验系统的测试表明,本文所设计的视频检索系统框架对于使用标准 SVM 处理视频流程后的查准率平均有 9.2% 以上的提升。

2 基于 SVM 的视频检索系统框架描述

视频检索系统的首先需要对视频进行特征的提取。特征提取一般都是先将视频根据镜头的切换分割成各个场景,然后提取出各个场景片段中的首尾帧和中间帧,这些帧就称之为关键帧,将这些关键帧的一些图像特征,比如颜色直方图,纹理信息,形状特征等提取出来,即可以作为视频的底层特征。

在完成底层特征的提取工作以后,则可以将底层特征通过一系列模式识别的算法将其转化为高级语义,如足球比赛,会议等,最后将所得到的语义对视频进行自动标注。用户查询时,则可以通过输入文本进行视频的直接检索。

本文的视频检索框架的流程如图 1 所示。

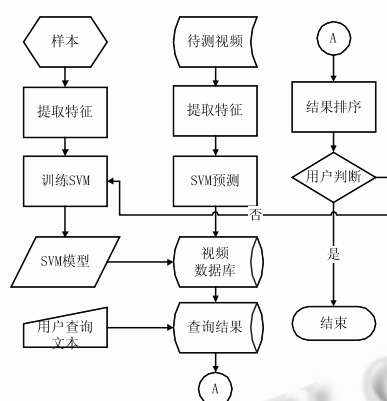


图 1 基于相关反馈和结果排序的 SVM 流程

本文提出的视频检索模型以 SVM 作为视频检索的核心部分,结合相关反馈,BOOST 思想和结果排序等多种方法围绕 SVM 提出了新的视频检索模型框架。本文通过双阈值算法对关键帧进行提取,进行视频底层特征的提取之后,进入图 1 中虚线框的部分,即本文章节 3 中所提出的基于反馈和结果排序的 SVM 算法。

3 针对视频检索优化的 SVM 算法

3.1 支持向量机(SVM)

支持向量机简称 SVM^[3],是统计学习理论中最年

轻的内容,也是最实用的部分。常用于模式识别当中。通常假设线性的问题中,所有数据点是 n 维实空间中的点。我们希望能够把这些点通过一个 $n-1$ 维的超平面分开。通常这个被

称为线性分类器。有很多分类器都符合这个要求。而其中使得属于两个不同类的数据点间隔最大的那个面,该面亦称为广义最优分割面。如果我们能够找到这个面,那么这个分类器就称为最大间隔分类器。

支持向量机是从广义最优分割面的基础上发展而来。对非线性问题,可以通过非线性变化将其转化到高维空间计算,得到最优分类面的高维空间后的计算实际上只需进行内积运算,而这种内积运算是可以用原空间中的函数实现的,我们甚至没有必要知道变换的形式。根据泛函的有关理论,只要一种核函数 $K(x_i, x_j)$ 满足 Mercer 条件,它就对应某一变换空间中的内积。因此采用适当的内积函数 $K(x_i, x_j)$ 就可以实现某一非线性变换后的线性分类,而计算复杂度却没有增加,此时分类函数变为:

$$f(x) = \text{sgn}\{\sum_{i=1}^l \alpha_i^* y_i K(x_i, x) + b^*\}$$

这就是支持向量机。概括地说,支持向量机就是首先通过用内积函数定义的非线性变换将输入空间变换到一个高维空间,在这个空间中求(广义)最优分类面。SVM 分类函数形式上类似于一个神经网络,输出是中间节点的线性组合,每个中间节点对应一个支持向量。

3.2 问题描述

假设原始样本集为 $U = (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 为样本集内容,根据关键帧是否属于语义设置。根据 U 训练 SVM 分类器为 D_1 , 当分类器 $D_1(x_k) > 0.5$ 时标志 $y_k = 1$ 。每次根据反馈的所训练的分类器为 D_k , 假设第 k 次用户反馈的样本集为 L_k , 第 k 个分类器训练的样本集 S_k , 当第 k 次反馈结束,则分类器 D_k 训练的样本为 S_k , 并更新原始样本集 U , 使得训练的样本为: $U = U \cup L_k$

3.3 基于相关反馈和结果排序的 SVM 算法

视频检索系统的机制决定了其可以有用户参与的过程,本文结合视频查询系统结果本身的特点,对基于反馈的 SVM 进行了改进。本文结合 BOOST 算法的思想,在 SVM 中引入了权重的概念。本文中与 BOOST 算法不同的是,不是通过调整弱分类器的权重生成最终的分器,而是根据每个 SVM 分类器的计算偏差,调整 SVM 分类器的权重从而生成最终的分器。本文

将根据反馈所得到的样本所训练得到的分类器保存为 D_k ，然后通过计算每个 D_k 的 y_i 偏差，赋予 D_k 不同的权重并最后构成最终分类器。

同时根据视频检索的特点，用户的正反馈有着一定的相同点，而负反馈则没有明确的相同点，所以本文调节了除了分类器 D_1 以外弱分类器的样本集，在样本集中加入了所有的正样本，而负样本则采用反馈中的结果，尽可能的取得正反馈的共同点。因此样本集 S_k 为：

$$S_k = L_k \cup S_{k-1} \cup \{X : X \in U, y = 1\}$$

BEGIN

(1) 始化样本集 U ，使得 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

中对于是否属于此标签设置 $y_i = 0, 1$

(2) 初始化标准分类器偏差 $\varepsilon_1 \leftarrow \sum_{i=1}^n |D_1(x_i) - y_i|$

(3) 训练分类器

根据 S_k 训练 D_k

$$S_k = L_k \cup S_{k-1} \cup \{X : X \in U, y = 1\}$$

更新分类器权重：

$$w_k = \frac{\varepsilon_k}{\sum_{i=1}^k \varepsilon_k}$$

最终分类器

$$C = \sum_{i=1}^k w_k \cdot D_i(x_m)$$

(4) 对结果进行排序(参见本文结果排序算法)；

(5) IF 用户不满意结果 THEN

 计算 $S_k = L_k \cup S_{k-1} \cup \{X : X \in U, y = 1\}$ ；

 回到 步骤 (3)；

ELSE

 结束；

ENDIF

END

同时 Xinmei Tian, Linjun Yang 等人的工作^[4]表明，对搜索结果进行朴素贝叶斯排序可以提高视频检索中前面结果的有效性。由于本文采用了基于 SVM 的分类，所以本文结合了 SVM 算法和结果排序，设计了基于 SVM 的视频结果排序，算法中利用超平面到向量的距离作为依据进行排序，采用在 3.2 章节中对 SVM 训练结束后所获得的超平面方程，计算关键帧中的特征到超平面的距离，设为超平面方程，则取距离，距离超平面最远的作为优先级最高最先显示。

由于 SVM 相较于朴素贝叶斯有着分类精度高，分类速度与训练样本个数无关等优点，并且由于本文中的特征分析中已经将基于 SVM 的结果排序所需要的数据计算完毕，所以采用基于 SVM 的结果排序，不仅在精度上，效率上都存在着较大的提升。

所以本文在反馈过程中加入了结果排序的流程。

算法描述：

BEGIN

(1) 根据 SVM 训练模型初始化方程 $wx + b = 0$

(2) 根据 $\frac{|wx + b|}{\|w\|}$ 计算出查询所得的视频到分割平面的距离

(3) 根据计算出的距离对查询的结果进行排序

END

结果排序可以进一步的调整弱分类器的样本集，由于结果排序的依据是数据到 SVM 模型中超平面的距离，所以在采用排序后前部的结果作为 SVM 的样本可以使得分类器的效果得到很好的提升。

4 结果与分析

通过采用 VC2008.net+OpenCV，构建了基于本文框架的实验系统，并改进了 LIBSVM 库函数对于 SVM 进行训练预测。

视频检测系统常用的分析数据时查准率和查全率，查准率表示查询返回数据中正确的数据占返回数据的比例，而查全率代表查询返回中所有正确的数据占有所有正确数据的比例。在采用同样的算法下对于传统的视频检索框架和采用了本文的视频检索框架的系统用 TREC2007 的视频进行查准率，查全率分析。结果表明，使用了本文提出的新的框架的系统的查准率和查全率得到了有效提高，以返回 10 个结果为例，具体数据见图 2。

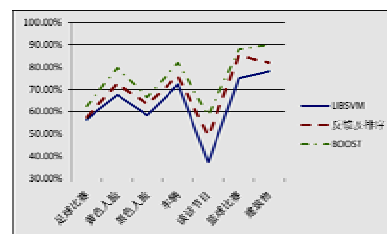


图 2 查准率分析

采用 TREC2007 中 200 个视频片段进行查准率提升分析，图 2 是使用 LIBSVM 直接训练后直接预测

的结果,加入了反馈和排序机制,以及加入 BOOST 权重后 SVM 预测的结果。采用公式:查准率=正确匹配的视频数/查询到的总视频数。



图 3 查询界面

实验结果表明本文反馈和排序在 SVM 的基础上结合后,所提出的基于 SVM 的视频检索系统框架对于不同语义的查询均能有大幅度的提升,尤其是不采用本文框架时查询效果较差的语义概念。

与正常直接采用 LIBSVM 中训练预测的结果相比较,本文框架后所构建的试验系统对于所有概念的查准率平均有 9%以上的提升。

5 结语与展望

本文展开了对视频检索系统框架设计的讨论,对 SVM 算法进行了针对视频检索系统的改进和应用。

本文中首先针对视频检索系统的特点,对于 SVM 引入了相关反馈与基于结果的排序的机制,其后引入了 Boost 算法中权重的概念,实验证明,相关反馈和

基于结果的排序以及 BOOST 算法权重的加入均能有效提高视频系统的检索能力,尤其是在不改进情况下查准,查全率较低的关键词提高较为明显,因此说明改进的 SVM 算法流程能更为有效的适用于各种概念的查询。

在其后的工作中,会对 SVM 反馈流程中的样本集的选取进行进一步的调整研究,同时在 SVM 核函数的选择上进行一定的研究,采用新的核函数;在条件成熟的情况下,本文描述的框架可以尝试将所得到的基础语义建立一个语义网络,根据推理机实现一定程度的对更高级语义的查询。

参考文献

- 1 Zhang L, Lin FZ, Zhang B. Support vector machine learning for image retrieval Proc. of the 2001 International Conference on Image Processing: ICIP 2001. Washington, DC: IEEE Computer Society, 2001:721 - 724.
- 2 Zavesky E, Chan SF Columbia University's Semantic Video Search Engine 2008, CIVR'08, July 7-9, 2008:545.
- 3 张学工.关于统计学习理论与支持向量机.自动化学报, 2000,26(1):32 - 41.
- 4 Tian XM, Yang LJ, Wang JD, Yang YC, Wu XQ, Hua Xb-Sg. Bayesian Video Search Reranking. MM'08, October 26-31, 2008.131 - 140.