

摘 要

近年来,随着视频压缩技术,硬件存储设备和网络传输技术的快速发展,视频、音频、图像等多媒体数据作为一种动态、直观和形象的数字媒体,承载了大量丰富的语义信息,广泛应用于各类信息服务和应用场合。

音频是多媒体信息的重要组成部分,是多媒体内容分析和理解的有效线索。鉴于原始音频数据本身是一种非语义符号表示和非结构化的二进制流,缺乏内容语义的描述和结构化的组织,给音频信息的处理和分析带来很大困难,因此如何自动而高效地从海量音频数据中提取音频中的结构化信息和高层语义是音频信息深度处理、基于内容的检索和辅助视频分析等工作的关键问题。基于内容的音频分类技术是解决这一问题的关键技术,是音视频内容分析、理解和检索的基础。

本文的工作围绕着基于内容的音频分类技术而展开,并针对体育视频中的音频分类问题,即音频关键字检测问题进行关键技术研究。研究工作涵盖了音频的层次化结构分析、底层音频特征的分析 and 提取以及基于统计理论的分类器设计。本文取得的研究成果如下:

1. 提出了基于 AdaBoost 模型的音频分类方法。其主要思想是通过自适应地组合弱分类器以得到增强分类器。与其它统计学习方法相比,这种方法准确有效、计算复杂度小,具有很好的泛化能力,并且可以在训练过程中自动选择分类效果更为有效的特征。
2. 针对类别不均衡的情况,提出了一种集成的自提升欠采样策略。其核心思想是对训练集中的大类样本进行多次欠采样,在相对均衡的样本集上训练弱分类器,最后合并多个弱分类器,得到总体分类器。采样的过程是有监督的,借鉴 AdaBoost 的思想指导采样的过程。通过这种方式,可以降低类不均衡问题所带来的不利影响,同时最大程度得保留大类样本中的有用信息。
3. 针对跨领域音频分类问题和同领域多类别分类问题,研究了多任务学习方法以及在音频分类中的应用。此方法的核心思想是从源域的未标注数据集中学习到最富有预测能力的低维映射特征向量,再将其应用于目标域的分类问题。

这种方法的优点在于可以深度挖掘大量未标注数据的有用信息，并且可以解决跨领域的音频分类问题，具有一定的通用性。

综上所述，本文研究了音频分类的核心问题，并针对体育视频自身的特点，提出了有效的音频关键字检测算法。

关键词：音频分类 AdaBoost 类不平衡学习 多任务学习 音频关键字检测

Technology Research of Audio Classification in Sports Video Analysis

Cai Shaojie

Supervised by: **Huang Qingming** (Professor)

With the rapid increase in speed and capacity of computer, hardware and networks, multimedia data, which includes image, video and audio, have been widely used in various applications.

Audio data is an integral part of multimedia data and effective clue to multimedia analysis and understanding. However, raw audio data is non-semantic and unstructured binary stream which lacks for semantic description and structured organization, bringing much difficulty for audio information processing and analysis. Therefore, how to automatically and effectively extract the structured information and high-level semantics is critical to audio and video analysis, understanding and retrieval. Content-based audio classification is the key technology to solve this problem.

The analysis framework is based on the content-based audio classification technology and taking sports video as the research object. The research consists of the audio structure analysis, low-level feature extraction and selection, statistics based audio classification algorithm. The main contribution of the thesis can be concluded as follows:

1. An AdaBoost- based audio classification algorithm is proposed. The main idea is combining multiple weak classifiers into a single strong classifier. Compared with other statistical learning method, this method is fast and accurate, and it can automatically select the most powerful features during training process.
2. Based on the class imbalanced dataset, a novel ensemble under-sampling strategy is proposed, which samples several subsets from the majority class, trains a learner using each of them, and combines the outputs of those learners. During the sampling process, each trained classifier is used to guide the training of subsequent classifiers. In this way, it can weaken the adverse

effect of the imbalanced dataset and retains the most informative data in the major class.

3. Multi-task learning is researched and applied to audio classification. The main idea is to learn a low-dimensional predictive structure from unlabeled data of the source domain and transfer the structure to the task domain. This method can deeply mine the useful information of the unlabeled data irrespective of the amount of the labeled data available, and it can be employed to tackle the audio classification problem across different domains.

In conclusion, the thesis has conducted a comprehensive research on audio classification and proposes effect audio keyword extraction algorithms considering the characteristics of the sports video.

Keywords: Audio classification, AdaBoost, class imbalance learning, multi-task learning, audio keyword extraction.