

Chapter 22

Statistical Machine Learning Approaches for Sports Video Mining Using Hidden Markov Models

Guoliang Fan

Oklahoma State University, USA

Yi Ding

Oklahoma State University, USA

ABSTRACT

This chapter summarizes the authors' recent research on the hidden Markov model (HMM)-based machine learning approaches to sports video mining. They will advocate the concept of semantic space that provides explicit semantic modeling at three levels, high-level semantics, mid-level semantic structures, and low-level visual features. Sports video mining is formulated as two related statistical inference problems. One is from low-level features to mid-level semantic structures, and the other is from mid-level semantic structures to high-level semantics. The authors assume that a sport video is composed of a series of consecutive play shots each of which contains variable-length frames and can be labelled with certain mid-level semantic structures. In this chapter, the authors present several HMM-based approaches to the first inference problem where the hidden states are directly associated with mid-level semantic structures at the shot level and observations are the visual features extracted from frames in a shot. Specifically, they will address three technical issues about HMMs: (1) how to enhance the observation model to handle variable-length frame-wise observations; (2) how to capture the interaction between multiple semantic structures in order to improve the overall mining performance; (3) how to optimize the model structure and to learn model parameters simultaneously. This work is the first step toward the authors' long-term goal that is to develop a general sports video mining framework with explicit semantic modeling and direct semantic computing.

DOI: 10.4018/978-1-60566-766-9.ch022

INTRODUCTION

Video mining is to discover knowledge, structures, patterns and events of interest in the video data, and its benefits range from efficient browsing and summarization of video content to facilitating video access and retrieval in a large database or online multimedia repository. There are various types of video data. According to different production and edition styles, videos can be classified into two main categories: *scripted* and *non-scripted* videos (Xiong, Zhou, Tian, Rui, & Huang, 2006). Scripted videos are produced according to a certain script or plan that are later edited, compiled, and distributed to users for consumption. News and movies are the examples of scripted videos. Most research efforts on scripted videos focus on the development of a Table-of-Content (TOC) that provides an overview of the video's naturally organized content to support efficient browsing and indexing. On the other hand, the events in non-scripted videos happen spontaneously and usually in a relatively fixed setting, such as meeting videos, sports videos, and surveillance videos. The research on non-scripted videos involves detecting high-lights and events-of-interest. In this chapter sports video mining is studied using American football as the case study.

Although sports videos are non-scripted content, there are still definite or repetitive structures and patterns. Using these structures and patterns, we can develop some flexible and effective tools for video browsing/indexing. Currently, there are two kinds of approaches for sports video mining, *structure-based* (Kokaram, et al., 2006; Xie, Chang, Divakaran, & Sun, 2003) and *event-based* (Assfalg, Bertini, Colombo, Bimbo, & Nunziati, 2003; T. Wang, et al., 2006). The former uses either supervised or unsupervised learning to recognize some basic semantic structures (such as canonical view in a baseball game or play/break in a soccer game) that can serve as an intermediate representation supporting semantics-oriented video retrieval, but usually cannot deliver high-level semantics directly. The latter one provides a better understanding of the video content by detecting and extracting the events-of-interest or the highlights, which could be very specific and task-dependent and requires sufficient and diverse training examples. Because these two approaches are complementary in nature, researchers have investigated how to integrate both of them in one computational framework. For example, a mid-level representation framework was proposed for semantic sports video analysis involving both temporal structures and events hierarchy (L. Y. Duan, Xu, Chua, Q. Tian, & Xu, 2003) and a mosaic-based generic scene representation was developed from video shots and used to mine both events and structures (Mei, Ma, Zhou, Ma, & Zhang, 2005).

Machine learning is one of the most feasible approaches for semantic video analysis (Kokaram, et al., 2006). Hidden Markov models (HMMs) have been the most popular tool in this area (Xie, Chang, Divakaran, & Sun, 2003). Our goal is try to establish a HMM-based machine learning framework that supports both structure analysis and event analysis and delivers the rudimentary building blocks for high-level semantic analysis. There are two major distinctions in our research. One is the concept of semantic space which supports explicit semantic modeling and specifies both mid-level and high-level semantic structures as well as relevant visual features, so that the task is formulated as two inference problems. The other is a new HMM that offers multi-faceted advantages for semantic video analysis.

In our recent research (Ding & Fan, 2006, 2007a, 2007b, 2008), we have discussed several HMM-based approaches that are able to support unsupervised learning for semantic modeling and computing. We assume that an American football video is composed of a series of consecutive play shots with all breaks and commercials removed. We are interested in two mid-level semantic structures that include “*play types*” (what happened) and “*camera views*” (where it happened), both of which are defined at the shot-level and can be specified by a set of frame-wise visual features in a shot. Although these seman-

23 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/statistical-machine-learning-approaches-sports/36999

Related Content

Machine Learning Through Data Mining

Diego Liberati (2012). *Machine Learning: Concepts, Methodologies, Tools and Applications* (pp. 23-31).

www.irma-international.org/chapter/machine-learning-through-data-mining/56128

A Proficient Hybrid Framework for Image Retrieval

Rajkumar Soundrapandiyanand Ramani Selvanambi (2021). *Applications of Artificial Intelligence for Smart Technology* (pp. 206-219).

www.irma-international.org/chapter/a-proficient-hybrid-framework-for-image-retrieval/265587

Developing Concept Enriched Models for Big Data Processing Within the Medical Domain

Akhil Gudivada, James Philipsand Nasseh Tabrizi (2020). *International Journal of Software Science and Computational Intelligence* (pp. 55-71).

www.irma-international.org/article/developing-concept-enriched-models-for-big-data-processing-within-the-medical-domain/258866

P3 Process for Object-Relational Data Migration to NoSQL Document-Oriented Datastore

Aicha Aggouneand Mohamed Sofiane Namoune (2022). *International Journal of Software Science and Computational Intelligence* (pp. 1-20).

www.irma-international.org/article/p3-process-for-object-relational-data-migration-to-nosql-document-oriented-datastore/309994

Soft Computing Based Statistical Time Series Analysis, Characterization of Chaos Theory, and Theory of Fractals

Mofazzal H. Khondekar, Dipendra N. Ghosh, Koushik Ghoshand Anup Kumar Bhattacharya (2013).

Handbook of Research on Computational Intelligence for Engineering, Science, and Business (pp. 30-61).

www.irma-international.org/chapter/soft-computing-based-statistical-time/72488