

Abstract

This thesis explores the problem of modeling dynamic events and structures based on a novel computational model, named an “interval-based hybrid dynamical system”. The model integrates two types of systems that have different concepts of time: dynamical systems, which are suitable for describing physical phenomena (consider time as physical metric entity), and discrete-event systems, which are suitable for describing human subjective or intellectual activities (consider time as ordinal state transition).

Firstly, we assume that a complex dynamic event such as human behavior consists of dynamic primitives (such as "open", "close", and "remain closed" in lip motions). Once the set of dynamic primitives is determined, a complex behavior can be partitioned into “temporal intervals” based on the primitives. Secondly, we assume that not only temporal orders but the duration lengths or temporal differences among beginning and ending time points of the temporal intervals, which we refer to as “timing structures”, have crucial information to understand dynamic events appear in human communication.

Based on the assumptions above, we propose an interval-based hybrid dynamical system, which has a two-layer architecture that consists of a finite state automaton and a set of linear dynamical systems. In this architecture, each dynamical system represents a dynamic primitive that corresponds to a discrete state of the automaton; meanwhile the automaton controls the activation timing of the dynamical systems. Thus, the overall system can generate, analyze, and describe complex dynamic events based on the structures of temporal intervals.

In spite of the flexibility of the systems, the learning process has a difficulty due to its paradoxical nature; that is, it requires us to solve temporal segmentation and system identification problems simultaneously. We therefore propose a two-step learning method. The first step of the method estimates the number of linear dynamical systems and its parameters based on the hierarchical clustering of dynamical systems, and the second step refines overall system parameters.

Experiments on simulated and real image data show that the proposed method successfully solves segmentation and system identification problems from input time-varying signals.

Applying the proposed model to describe structured dynamic events that consists of multipart primitives, we can extract and analyze dynamic features based on the timing structures extracted from temporal intervals. We examined the effectiveness of using the timing structures to analyze and discriminate fine-grained facial expression categories such as intentional and spontaneous smiles of which existing methods had difficulty to represent the difference.

Finally, we propose a “timing structure model” that directly represents timing structures in multimedia signals, such as synchronization and mutual dependency with organized temporal differences among temporal patterns of media signals. Experiments on simultaneously captured audio and video data show that time-varying signals of one media signal can be generated from another related media signal by using the trained timing structure model.