博士論文(要約)

# An Information-Theoretic Approach to Hierarchical Change Detection (階層的変化検知への情報理論的アプローチ)

Shintaro Fukushima (福島 真太朗)

Abstract

With recent advances in technology, it has become even more important to discover valuable information from data. In this study, we address the issue of extracting deep knowledge of changes in data streams, in addition to when changes occur. The knowledge includes how frequently and how significantly changes occur, how long changes continue, and where changes come from. There have been only a few studies that addressed this topic. To formulate these issues from a unified perspective, we propose a new concept of hierarchical change detection, and define the following three detection tasks: 1) metachange detection (detection of changes in patterns of changes). 2) change duration detection (detection of durations of changes). 3) hierarchical change detection in latent variable models (detection of changes in different layers of latent variable models). Our research questions include how to characterize and detect the changes, how to integrate or decompose them, and how to represent tradeoffs between them. We solve these issues with information-theoretic approaches: encoding magnitudes of changes from viewpoints of the minimum description length principle (MDL) and robust and adaptive learning with online learning.

First, for the metachange detection, we consider how to detect changes in patterns of how frequently and how significantly changes occur. We propose a new notion of metachanges, and define associated metachange statistics with MDL. The key idea of the metachange detection is to learn patterns of a sequence of change points, and then detect these changes. Second, in regard to the change duration task, we address the issue of how to discriminate outliers and essential changes in underlying data generating mechanisms. There is an obvious tradeoff between mitigating or removing outliers and adapting to the essential changes. We propose an online learning algorithm to quantitatively evaluate the tradeoffs. Finally, for the hierarchical change detection in latent variable models, we consider how to detect changes in different layers of latent variable models in a hierarchical way. We propose an information-theoretic framework for detecting changes of different levels of latent variable models from a viewpoint of MDL.

The three tasks of hierarchical change detection and their information-theoretic approaches open a new door to discovering change characteristics in data streams beyond conventional change detection.