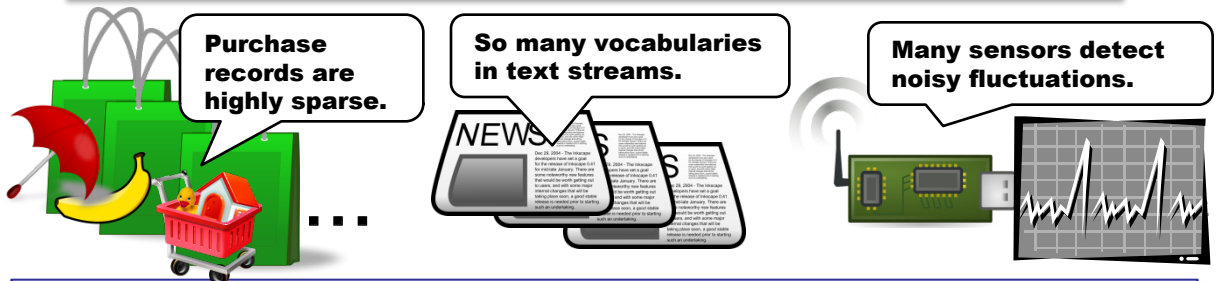# 1 Understanding stream data from essential elements
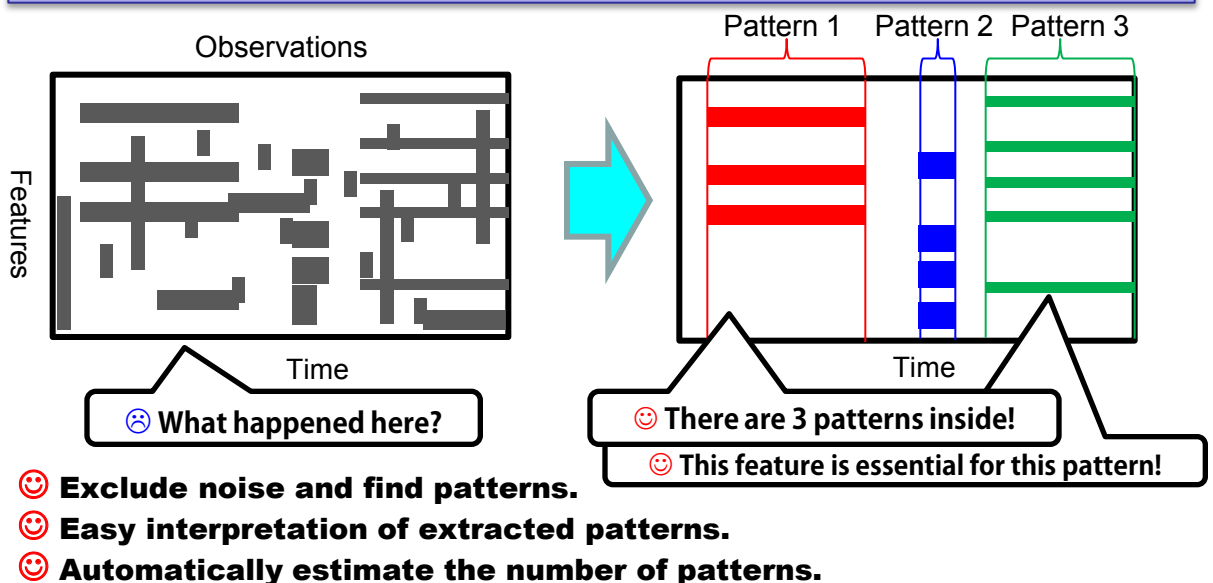
Feature selection for high-dimensional time series data

Abstract— In this research, we develop a pattern mining technique from high-dimensional time series data. Given the time series data, our technique finds hidden segments of time steps called "states" or "patterns". Furthermore, our technique automatically extracts "essential" features that characterize the patterns. The technique is especially suitable for the analysis of time series data containing many noisy features because it automatically excludes these noisy features, which are not responsible for pattern changes. The technique will also reduce the number of man-hours needed to analyze the time series data because it selects the essential features of time series patterns to interpret the analysis results.

Are all features in the time series informative and essential?

Purchase records are highly sparse.

So many vocabularies in text streams.

Many sensors detect noisy fluctuations.

NEWS

. . .

Our technique extracts "essential" features that characterize patterns within high-dimensional observation streams.

Observations

Features

Time

☹ What happened here?

Pattern 1 Pattern 2 Pattern 3

Time

☺ There are 3 patterns inside!

☺ This feature is essential for this pattern!

☺ **Exclude noise and find patterns.**
☺ **Easy interpretation of extracted patterns.**
☺ **Automatically estimate the number of patterns.**

Related works

[1] K. Ishiguro, N. Ueda and H. Sawada. "Dynamic Subset Clustering," in *Proc. IBIS*, 2011.

Contact

Katsuhiko Ishiguro   Learning and Intelligent Systems Research Group, Innovative Communication Laboratory
E-mail : ishiguro.katsuhiko{at}lab.ntt.co.jp (Please replace {at} with @)