

## データストリームにおける遅延相関の検出

櫻井 保志

NTT サーバースペース研究所  
E-mail: sakurai.yasushi@lab.ntt.co.jp

あらまし 近年、データストリームの処理やマイニングに関する研究が盛んである。ネットワーク分析、センサーモニタリング、金融データ分析、ムービングオブジェクトのトラッキングなど、データストリームは様々な分野で用いられている。本講演では、BRAIDと呼ぶデータストリーム間の遅延相関検出のための手法について述べる。BRAIDは、半無限長のデータストリームを、インクリメンタル、高速、そして省メモリで処理することができる。さらに、BRAIDに関する理論的な分析についても述べる。これはサンプリング理論に基づいたものであり、BRAIDの精度について示したものである。実データと人工データを用いた実験では、相対的な誤差が最大で1%であり、精度良く遅延を検出している。計算速度については、厳密に計算する場合と比べ最大で40,000倍の高速化を達成している。

キーワード データストリーム処理、パターン検出、遅延相関

## Detecting Lag Correlations in Data Streams

Yasushi Sakurai

NTT Cyber Space laboratories  
E-mail: sakurai.yasushi@lab.ntt.co.jp

**Abstract** The processing and mining of data streams have attracted on increasing amount of interest recently. Data streams appear in a variety of settings, such as environmental, medical and socioeconomic systems. Typical data-stream applications include network analysis, sensor monitoring, financial data analysis, and moving object tracking. In this talk, I will introduce BRAID, a method to detect lag correlations between data streams. BRAID can handle data streams of semi-infinite length, incrementally, quickly, and with small resource consumption. I will also introduce a theoretical analysis, which, based on Nyquist's sampling theorem, shows that BRAID can estimate lag correlations with little, and often with no error at all. Our experiments on real and realistic data show that BRAID detects the correct lag perfectly most of the time (the largest relative error was about 1%); while it is up to 40,000 times faster than the naive implementation.

**Keywords** data stream processing, pattern discovery, lag correlations