



Title	Streaming Algorithms for Dynamic Pattern Mining
Author(s)	川畑, 光希
Citation	大阪大学, 2021, 博士論文
Version Type	
URL	<a href="https://hdl.handle.net/11094/82289">https://hdl.handle.net/11094/82289</a>
rights	
Note	やむを得ない事由があると学位審査研究科が承認したため、全文に代えてその内容の要約を公開しています。全文のご利用をご希望の場合は、<a href="https://www.library.osaka-u.ac.jp/thesis/#closed">大阪大学の博士論文について</a>をご参照ください。

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

## 論文内容の要旨

氏 名 ( 川畑 光希 )

## 論文題名

Streaming Algorithms for Dynamic Pattern Mining  
(時系列パターンマイニングのためのストリームアルゴリズムに関する研究)

## 論文内容の要旨

There naturally exists time series data in wide range of domains, such as in economics, biology and industry, because we can observe data with timestamps that regularly pass by to analyze and discover valuable insights. Dynamics in such data have a lot of importance. A climate monitoring, for example, we are interested in temperature measurements at every hour or every day to make decisions. In industrial contexts, we want to reveal reasonable causes in machine troubles and defective products by monitoring sensor measurements. Although mining methods of time series have been well studied for last decades and been providing solutions satisfying these demands, we have faced more complex problems represented by three vital considerations: (a) handling larger, unlimited data samples, (b) complexity of model design and (c) interpretability of analytical results, due to the advent of big data era, in which we can access a massive amount of data and therefore the ideal methods should satisfy these concepts.

This thesis discusses streaming algorithms that is useful, effective and practical approaches to analyze a large flow of time series, which is known as data streams. To address the above three challenges, the algorithms should enable discovery of important sequential patterns underlying data and help us understand generating mechanisms or forecast upcoming situation. The first proposed algorithm particularly focuses on mining multi-dimensional data streams, and how to know important patterns and its total numbers automatically, based on a statistical model and a theory in data compression and summarization. The second algorithm tackles mining high-order time series, so-called tensor. The algorithm is designed so that it can decompose a given tensor stream into several latent factors that explains non-linear dynamics and seasonal effects. The third algorithm incorporates a non-linear equation for epidemic modeling with a stream forecasting framework, which enables us to share and exploit obtained models that varying over time with arbitrary locations.

Throughout extensive experiments using real world datasets that support the effectiveness of the three approaches, we will learn why streaming algorithms are important and what we can find out in data streams. The results also show that the proposed methods can handle large data streams in an efficient ways with keeping high performance compared with their baselines and state-of-the-art methods. Finally, the brief discussion about future directions is provided, containing extension of the proposed methods.

## 論文審査の結果の要旨及び担当者

氏 名 ( 川畑 光希 )			
	(職)	氏 名	
論文審査担当者	主 査	教授	櫻井 保志
	副 査	教授	橋本 昌宜
	副 査	教授	鬼塚 真
	副 査	准教授	松原 靖子

## 論文審査の結果の要旨

時系列パターン解析は、ユーザ行動の理解や将来予測を行うための要素技術として広く利用されているが、大規模データの解析においては、すべてのデータを保持することなく、効率的に解析を行うストリームアルゴリズムが必要とされる。本論文は、大規模時系列データストリーム環境に適用可能な時系列パターン検出、および、時系列パターンに基づく将来予測アルゴリズムの開発に取り組んだものである。本論文では、実データを用いた評価実験を実施し、以下の成果を得ている。

## [自動パターン検出のためのストリームアルゴリズムの開発]

IoTの普及に伴い複雑化する多次元データストリームの解析では、データに関する高度な事前知識を必要とせず、自動的に時系列パターンを抽出することが望ましい。本研究では、情報理論に基づく符号化スキームを応用し、データストリームを構成する時系列パターンの種類と変化点を自動的、かつリアルタイムに検出するためのアルゴリズムを開発した。提案手法は既存のオフライン手法と比較して同等の精度を示しながら、高速に動作することを明らかにした。

## [時系列テンソルのためのストリームアルゴリズムの開発]

オンライン活動履歴などの大規模テンソルデータは、多角的な観点から時系列パターンを分析し、データを構成する重要な要素を明らかにすることで、効果的な将来予測を行うことが可能となる。本研究では、非線形動的システムに基づく新たな時系列モデルを用いて、テンソルデータに含まれるトレンド、季節成分、および、類似パターンによる属性のグループ化を自動的に行うアルゴリズムを開発し、将来予測の精度を向上できることを明らかにした。

## [感染症データのためのストリームアルゴリズムの開発]

本研究では、急激に拡大する感染パターンを効果的に抽出するため、感染症数理モデルに基づく将来予測アルゴリズムを開発した。提案手法は、感染者数や回復者数などの統計値をテンソルデータとして扱い、各地域で推定された感染パターンを共有することで、高速に動作しながら、高い精度で将来の感染者数を予測することが可能であることを明らかにした。

以上の研究成果は、時間経過に伴い傾向が変化する大規模データストリームの解析において、効率的な時系列パターンの抽出を可能にし、それらの特徴が情報圧縮や将来予測の精度を向上させることを示した。これらのことから、本論文は、データストリーム解析の性能向上、適用範囲の拡大に大きく貢献するものであると期待できる。よって、博士（情報科学）の学位論文として価値のあるものと認める。