



Title	Causal inference on spatio-temporal point process data based on Granger Causality
Author(s)	Pavasant, Nat
Citation	大阪大学, 2024, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/96232
rights	
Note	

Osaka University Knowledge Archive : OUKA

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

Osaka University

Abstract of Thesis

Name (Nat Pavasant)	
Title	Causal inference on spatio-temporal point process data based on Granger Causality (グレングァー因果性に基づく時空間点過程データに対する因果推定)
<p>Abstract of Thesis</p> <p>Real-world data are usually in the form of spatio-temporal data, such as weather systems, transport demand, and disease outbreaks. However, the complexity of this type of data means that analysis techniques are not as well established. Identifying relationships, specifically causal relationships, within the spatio-temporal data can yield further understanding of natural phenomena, but the area is not well understood. This work proposed a method to extract causal relations of clusters from multi-dimensional event sequence data. The proposed Granger Cluster Sequence Mining (G-CSM) algorithm identifies the pairs of spatial data clusters that have causality over time with each other. It extended the Cluster Sequence Mining algorithm, which utilized a statistical inference technique to identify occurrence relation, with a causality inference based on Granger causality. In addition, the proposed method utilizes a false discovery rate to control the significance of the causality. The method was tested using both synthetic data and semi-real data and can extract embedded causal relations with high F-scores over different sets of data even under high spatial noise. False discovery rate also helps to increase the accuracy even more under such cases and also makes the algorithm less sensitive to the hyper-parameters. Furthermore, a local density estimation procedure was also proposed. This procedure is a pre-processing step to the vector autoregressive modeling of point-process data, a process which was used during Granger causality inference in the proposed algorithm, by applying a density estimation. Results on synthetic data showed that the procedure improved model accuracy, especially under sparse data.</p>	

論文審査の結果の要旨及び担当者

氏 名 (Nat Pavasant)			
	(職)	氏 名	
論文審査担当者	主 査	准教授	福井 健一
	副 査	教授	沼尾 正行
	副 査	教授	藤崎 泰正
	副 査	教授	谷田 純

論文審査の結果の要旨

本論文は、時空間上の点過程の観測データにおいて、時空間パターン抽出法に関する研究を行ったものである。このような時空間パターンの抽出は、例えば神経科学や地理情報科学において応用可能な技術である。本論文第3章では、従来の時空間パターン抽出法に対して、Granger因果性に基づく尤度関数を導入することで、Granger因果性の意味合いで因果性を持つ時空間パターン抽出法としてGranger Cluster Sequence Mining (G-CSM) を提案している。その際、多数の候補パターンに対するGranger因果性検定の誤検出を抑えるため、偽発見率 (False Discovery Rate; FDR) 制御による調整を導入した評価関数を提案している。実験では正解となる時空間パターンを埋め込んだ人工データに対して、提案法の各種ハイパーパラメータの影響、ノイズ耐性を綿密に検証している。従来法では全く検出が不可能であったノイズレベルにおいて、提案法は正解パターンの抽出に成功している。さらに、特徴空間として画像データを用いた半実データにおいても、従来法と比較して埋め込まれた正解パターンを高精度に抽出できることが示されている。

続く第4章では、Granger因果性の尤度のベースであるベクトル自己回帰 (Vector Autoregression; VAR) モデルにおいて、点過程データに適用する際の時間離散化の問題に対処する方法を提案している。ナイーブな区間による離散化は疎なデータに対して、入力系列長と離散化粒度の影響が大きくなる。本論文では、区間によるデータ点の数え上げを1次元密度推定に置き換えることで、VARモデルへの入力次元数を抑えつつ、系列長を伸ばす方法を提案している。その際、局所密度推定によって、計算コストを抑えると共にベクトル自己回帰モデルへの入力を可能にしている。特に提案法は疎なデータにおいて効果を発揮することが人工データを用いた実験により確認されている。

以上を要するに、本論文はGranger因果性に基づく時空間パターン抽出に関する研究を行い、その有効性を確かめたもので、時空間データ解析において情報科学技術の果たす役割の進展に大きく貢献するものである。よって、博士(工学)の学位論文として価値のあるものと認める。