

## BOCD を用いた変化点検知と予測 Change point detection and prediction using BOCD

下野将暉<sup>†</sup> 鈴木海友<sup>‡</sup> 小林正弘<sup>¶</sup> 松澤智史<sup>\*</sup>

Masaki Shimono Kaiyu Suzuki Masahiro Kobayashi Tomofumi Matsuzawa

### 1. はじめに

現代社会で、膨大な情報が蓄積されることで生成される時系列データは、一定の時点で統計的な性質が変化することがある。そのような時系列データに対する変化点検知は、気象環境の変化の検知 [1] や顔特徴量時系列データから注目点を抽出するために活用される [2] など、様々な分野で応用されている。データの生成プロセスや統計的特性が変化する時点という意味での変化点を検知する手法は複数存在しており、そのうちの一つに特異スペクトル解析に基づく手法がある [3]。しかしながら、当手法は計算コストが高く、データのリアルタイム処理に向かない特徴をもつ。一方で、Bayesian Online Changepoint Detection (BOCD) [4] というオンライン変化点検知手法が存在する。

### 2. 変化点検知手法

本研究では、変化点検知を行うために統計手法である BOCD に着目した。この手法を選択した理由は 2 つである。

1. 平均や分散をはじめとした、様々な統計的特性の変化を捉えることが期待される。
  2. オンラインで変化点を検知できる。
- 1 点目の利点について、統計的特性の代表である平均と分散を変化させた自動生成データを用いて、BOCD を適用させた予備実験の結果を図 1、図 2 に示す。

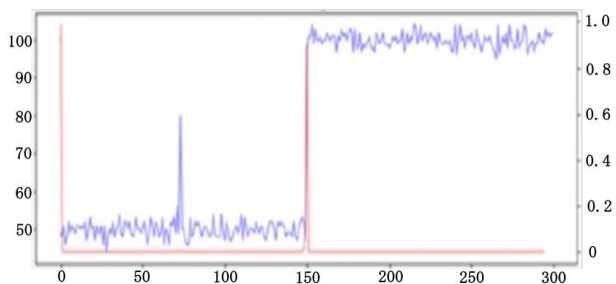


図 1 平均の変化に対する BOCD の動き

なお、図 1、図 2 における横軸は時刻を表す。青線は時系列データを表すグラフであり、赤線は変化点検知率を表すグ

<sup>†</sup> 東京理科大学大学院創域理工学研究科情報計算科学専攻 Department of Information and Computing Science, Graduate School of Science and Technology, Tokyo University of Science

<sup>‡</sup> 東京理科大学創域理工学部電気電子情報工学科 Department of Electrical, Electronics and Computer Engineering, Tokyo University of Science

<sup>¶</sup> 東海大学理学部情報数理学科 Department of Informatics and Mathematics, Faculty of Science, Tokai University

<sup>\*</sup> 東京理科大学創域理工学部情報計算科学科 Department of Information and Computing Science, Faculty of Science and Technology, Tokyo University of Science

ラフである。縦軸の左の値はデータ数、縦軸の右の値は変化点検知率を表す。図 1 において、時刻 70 でデータ数が急激に伸びているが、これはその時刻のみに現れた外れ値であり、変化点検知率は上昇していない。一方、時刻 150 で 50 から 100 に平均が変化しているが、これに対しては時刻 150 で変化点検知率が上昇し、その後は変化点検知率が上昇していない。

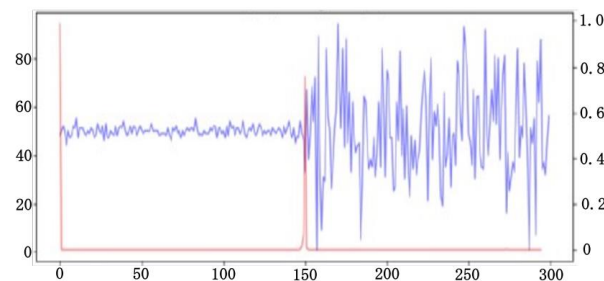


図 2 分散の変化に対する BOCD の動き

図 2 では、時刻 150 の前後で平均ではなく分散が変化している。時刻 150 で分散が 2 から 20 に変化しているが、時刻 150 のみで変化点検知率が上昇していることから、本研究の手法は分散の変化も捉えていることがわかる。また、2 点目の利点について、オンライン変化点検知モデルは 1 つのモデルに順次データを追加していく構造であるため、継続的に変化点を検知することができる。学習のたびに全てのデータを入力する必要のある、オフライン変化点検知モデル [5] [6] と比較したとき、計算量を抑えることができる。すなわち、オフライン変化点検知モデルでは実現できない、変化点を検知した際の早急な施策の立案が可能となる。

### 3. データの概要

本研究では 2 種類のデータを使用した。まず、政府統計ポータルサイトである e-Stat [7] の機械受注統計調査より、風水力機械、運搬機械、産業用ロボット、金属加工機械、化学機械、冷凍機械、合成樹脂加工機械、繊維機械、建設機械、鉱山機械、農林用機械、その他の産業機械の 12 種類の機種別機械受注額を活用した。そして、統計ダッシュボードの日経平均株価に関するデータを活用した。なお、これらは 2005 年 4 月から 2024 年 3 月までの月次データである。

### 4. BOCD による変化点検知の確認

本研究に活用した、日経平均株価データに対して BOCD を適用した結果を示す (図 3)。なお、BOCD は状態空間モデルを基礎としており、予備実験を行った結果、実装段階で変更可能な平滑化数を 10 と設定した。青線が日経平均株価の変移を表したグラフであり、赤線がその時点での変化点検知率を示している。縦軸の左側の数値は平均株価を表し、縦軸の右側の数値は変化点検知率を表している。BOCD は統計的性質の変化を捉える手法であり、株価の場合、急激な上昇及び下落に関しては時系列モデルの性質が変化したとして、

変化点と捉える。一方、株価の緩やかな上昇や下落は、実際のモデルが持っていた性質の可能性が高いため、統計的性質が変化したとは考えにくい。そのため、BOCD ではそれを変化点としては捉えない。以上の点を考慮すると、図 3 は株価の統計的性質の変化を検知できている。

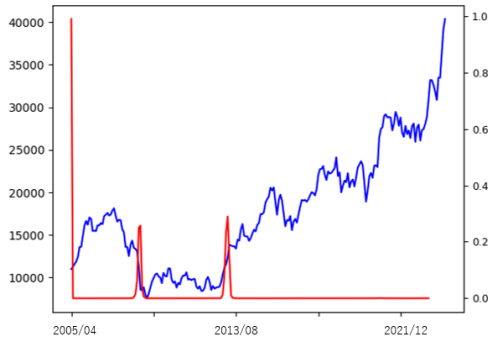


図 3 日経平均株価に BOCD を適用した場合

## 5. 変化点予測

図 3 の事例から明らかな通り、BOCD を活用することで、ある時点の日経平均株価データを入力したとき、平滑化数だけの時間を経たのちに、その時点が変化点を迎えているかを推定することができる。一方、小さな平滑化数を設定し、即座に警告を出すことを意図した場合、変化点の検知精度が下がる(図 4)。したがって、変化点に対する施策を事前に打つためには、変化点を予測することが必要となる。



図 4 平滑化数 1 の場合(日経平均株価)

また、合成樹脂加工機械の受注額データに対して BOCD を適用した結果を赤線と青線で示すと共に、図 3 における変化点確率の変移を黒線で示した(図 5)。受注額の増減を検知できており、さらに変化点確率が同時期に高まることわかる。合成樹脂加工機械以外の産業機械の受注額に関しても同様に、日経平均株価と変化点確率の高い時期が連動する場合があると推察したとき、機械学習を活用することで、日経平均株価の変化点確率を産業機械の受注額データを参考にして予測できると考えられる。

## 6. 今後の予定

Transformer[8]を活用し、週単位で地域別のインフルエンザ疾患患者比率を予測する先行研究[9]がある。また、過去のデータから有用な情報を抽出し、高精度な時系列予測を可能とする TimeFM[10]や Transformer と同程度以上の性能を出せることが示された xLSTM[11]というモデルがある。現時点でこれらを実装する上で必要な環境を仮想環境上に構築しており、今後は日経平均株価の変化点確率値を目的変数とすると共に、12 種類の機械受注額の変化点確率

の推移を説明変数と設定した上で、機械学習モデルを実装し、変化点予測を行う。また、多様な時系列データに対して変化点検知を行うことを目的として、BOCD に入力するのに適したデータの分布を明らかにすると共に、拡張を行う。

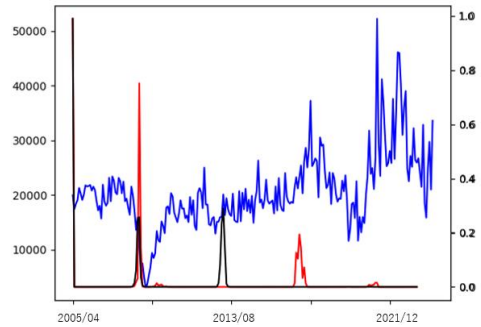


図 5 合成樹脂加工機械の受注額に BOCD を適用した場合と図 3 の重ね合わせ

## 7. まとめ

事前に用意した 2 種のデータを各々 BOCD に適用した結果、BOCD が変化点検知を行えることが確認された。また、活用する機械学習モデルを工夫し、変化点予測を行うと共に、BOCD の拡張についても取り組む。

### 参考文献

- [1] 前原宗太郎, 福井健一, 富田智彦, 小野智司, “気象時系列データにおける変化点検知の基礎検討”, 人工知能学会全国大会(第 31 回), 2017.
- [2] 平山孝輔, 佐伯幸郎, 中村匡秀, “変化点検知を活用した顔特徴量時系列データからの注目点の抽出”, 信学技報, vol.120, no.432, HCS2020-78, pp. 60-66, 2021.
- [3] Valentina Moskvina, Anatoly Zhigljavsky, “An Algorithm Based on Singular Spectrum Analysis for Change-Point Detection”, Communication in statistics-Simulation and Computation, 32(2), 319-352, 2011.
- [4] Ryan Prescott Adams, David J.C. MacKay, “Bayesian Online Change-point Detection”, arXiv:0710.3742, 2007.
- [5] Paul Fearnhead, “Exact and Efficient Bayesian Inference for Multiple Change-point problems”, Statistics and computing 16.2, pp.203-213, 2006.
- [6] Xuan Xiang, Kevin Murphy, “Modeling Changing Dependency Structure in Multivariate Time Series”, ICML, pp.1055-1062, 2007.
- [7] <https://www.e-stat.go.jp/dbview?sid=0003355226>  
閲覧日 2024 年 6 月 13 日
- [8] Vaswani, Ashish and Shazeer, Noam and Parmar, Niki and Uszkoreit, Jakob and Jones, Llion and Gomez, Aidan N and Kaiser, ukasz and Polosukhin, Illia, “Attention is all you need”, Advances in neural information processing systems, vol.30, 2017.
- [9] Neo Wu, Bradley Green, Xue Ben, Shawn O’Banion, “Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case”, arXiv:2001.08317, 2020.
- [10] Abhimanyu Das, Weihao Kong, Rajat Sen, Yichen Zhou, “A decoder-only foundation model for time-series forecasting”, arXiv:2310.10688.
- [11] Maximilian Beck, Korbinian Pöppel, Markus Spanring, Andreas Auer, Oleksandra Prudnikova, Michael Kopp, Günter Klambauer, Johannes Brandstetter, Sepp Hochreiter, “Extended Long Short-Term Memory”, arXiv:2405.04517.