

AI による数法則発見の時系列データへの 拡張と金融データへの応用

Modeling and Visualization of Social Reality
Using Latent Profile Analysis and Number Law Discovery Methods
for Evidence-Based Policy Making

蒲田 涼馬 (Ryoma Gamada)
u455007@st.pu-toyama.ac.jp

富山県立大学大学院 工学研究科 電子・情報工学専攻
情報基盤工学講座

N212, 09:30-10:00 Tuesday, February 13, 2024.

はじめに

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

数値実験並びに
考察

おわりに

はじめに

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

数値実験並びに
考察

おわりに

情報技術の発達により、社会における様々なデータを観測・収集することが可能に

→ 経済分析においても将来予測などの研究が急速に発展.
しかし要因分析に関する研究はそれほど進んでいるとは言えない.

経済に影響を与える要因を分析する研究

因果探索による要因分析.
シンボリック回帰を用いた要因分析.

本研究

シンボリック回帰を用いて分析を行う.

はじめに

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

数値実験並びに
考察

おわりに

なぜシンボリック回帰？

経済分野では、原因と結果の間に成り立つ関係性が重要

→ 複数の要因が複雑に影響しあうため、因果探索では具体的にどのような絡み合って影響を与えるかを詳細に分析できない。

アプローチ

公開されている金融データ、経済データ、市場間データを用いて分析を行い、データ間の関係性を数理モデルによって表す。

数理モデルの例

$$(\text{データ A}) = 2.0 \cdot (\text{データ B}) + 1.0 \cdot (\text{データ C}) - 1.0 \cdot (\text{データ D})$$

はじめに

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

数値実験並びに
考察

おわりに

公開されているデータ

XMMT5 や Investing.com, 日本銀行時系列サイトで様々なデータが公開されている.

Table 1: 公開されている様々な経済データ

データ項目	
為替レート	金利
コモディティ価格	エネルギー価格
マネーストック	ボラティリティ指数
出来高	スプレッド
株価指数	ニュース

はじめに

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

数値実験並びに
考察

おわりに

目的

時系列経済データを用いて，時系列を考慮した数法則の発見を行いデータ間の関係性をモデル化する手法を提案する．

はじめに

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

数値実験並びに
考察

おわりに

使用する手法の概要

機械学習を用いたシンボリック回帰手法である「End to end symbolic regression with Transformers」を拡張させ、時系列を考慮した分析を行う。可読性と得られる情報量を重視し、人間が式を見ることである程度その式が何を表しているのかをわかるレベルのものを生成させる。

はじめに

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

数値実験並びに
考察

おわりに

End to end Symbolic Regression with Transformers の概要

データから数式を自動発見する深層学習アプローチ

従来のシンボリック回帰の課題: 計算コストが非常に高い.

Transformer アプローチの着想:

数式は、演算子、定数、変数といった要素が並んだシーケンスとして表現する.

例) $y = x + 2 \cdot \sin(z) \rightarrow + \times * 2 \sin z$ Transformer はシーケンスデータの複雑なパターン学習と高速なシーケンス生成能力を持つため、数式発見に応用できる.

Transformer のメカニズム

Transformer はエンコーダとデコーダから構成され、Attention メカニズムがその核をなす。

■ Attention メカニズム

シーケンス内の各要素が、ほかのどの要素に注意を向けるべきかを動的に判断し、その重要度について重みづけを行う。

$$Scores = QK^T \quad (1)$$

$$ScaledScore = \frac{QK^T}{\sqrt{d_k}} \quad (2)$$

$$AttentionWeight = softmax(\frac{QK^T}{\sqrt{d_k}}) \quad (3)$$

$$Attention = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (4)$$

ここで Q はクエリ行列, K はキー行列, V は値, d_k はキーの次元を意味する。

はじめに

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

数値実験並びに
考察

おわりに

学習

- 入力データの埋め込み.
→ 入力データをエンコーダが処理可能なベクトルに変換する.
- エンコーダによるパターンの抽出
→ 入力を受け取り, Attention 機構を通してデータの関係性を学習する.
- デコーダによる数式の学習
→

数法則の発見

使用したデータ

項目：10 種類の経済データ

対象年：2015 年 1 月 1 日から 2024 年 12 月 31 日までの 10 年間

データ数：土日祝日を除いた日足で 2355 のデータ (前半 8 割で学習, 2 割でテスト)

目的変数：USDJPY の t のときの値

説明変数：Table2 の $t-1$ のときの値

Table 2: 数値実験に用いたデータ

データ項目	
SP500 価格	日経平均株価
日米の金利差 10 年	日本 10 年国債
米国 10 年国債	日米の金利差 2 年
オイル価格	金価格
USDJPY	VIX 指数

はじめに

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

数値実験並びに
考察

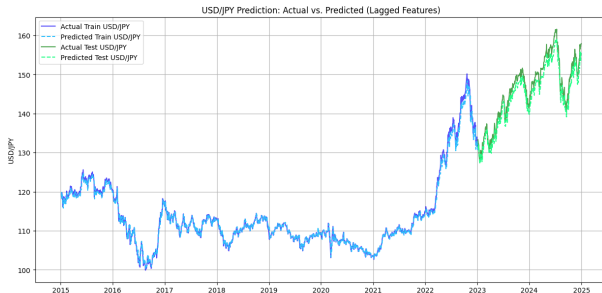
おわりに

実験結果

USDJPY を求めた結果を以下に示す.

決定係数

- トレーニングデータについての決定係数 : 0.9921
- トレーニングデータについての RMSE : 0.9921
- test データについての決定係数 : 0.9329
- test データについての RMSE : 2.0724



実験 1 の結果と考察

12/14

はじめに

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

数値実験並びに
考察

おわりに

$$\begin{aligned}
 & -0.10040714384915 \cdot x_6 - 0.000198358776965863 \cdot x_7 + 0.951613884352736 \cdot x_8 + 5.3976 \\
 & 0.011125679 + \\
 & -2.21694913680499e-5 \cdot x_9 - 0.000115127670935178 \cdot x_{10} + 0.043894457 \\
 & 2668756 \cdot x_{11} + 333.516649543171 \cdot x_{12} - 130.630846400977 \cdot x_{13} + (0.301880750900265 \\
 & 0.677 \\
 & - 0.00456788346207082 \cdot x_{14}) \cdot (1.85838014281174e-5 \cdot x_{15} + 11.5184087304255 \cdot x_{16} - (4 \\
 & .35176808096492 \cdot x_{17} + 6.64848083125045) \cdot (0.00108863330413424 \cdot x_{18} - 0.054506250 \\
 & 8133566) - 199.764377322288) + 249.687084647262
 \end{aligned}$$

$$0.952 \cdot USD / JPY_{\text{logit}} - 0.1 \cdot \text{金利差}_{JP-US10Y} +$$

$$0.677 \cdot \ln(333.517 \cdot JPY_{\text{logit}} - 130.631 \cdot US_{10Y_{\text{logit}}}) + 0.044 \cdot \text{金利差}_{JP-US2Y} + (0.302 - 0.005 \cdot WTI_{Oil}) \cdot (11.518 \cdot VIX - (0.001 \cdot WTI_{Oil} - 0.055) \cdot (4.352 \cdot \text{金利差}_{JP-US2Y} + 6.648) - 199.764) + 249.687084647262$$

はじめに

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

数値実験並びに
考察

おわりに

可読性の向上

可読性の高さを決める要因として以下が考えられている。

- ノード数, 深さ, 演算子の数と種類, 項の数, コード長

可読性向上についての研究

- 損失関数に複雑度を組み込んで複雑になりすぎず, 精度もある程度維持させる.
- 短い数式や簡潔な式に高いスコアをつけ, 学習させる.

はじめに

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

数値実験並びに
考察

おわりに

まとめ

手法を色々検討し、今後使っていく手法を決定した。また、精度と可読性を確認した。

- 決定係数は 0.93 とある程度高く、経済をモデル化できている可能性もある。
- 現状では可読性が低く、得られる情報量が少ない。

今後の展望

- まずは手法について完璧に理解する。
- 精度を維持しつつ可読性を向上させるためのシステムを考え、実装してみる。
- 今回は $t-1$ からの影響を考えたが、もっと色々な時間から受ける影響を考慮できるようにする。