

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

経済における波及
メカニズム

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

今回やったこと

データの拡張

実験計画法

実験計画法の手法

おわりに

本研究の背景 1

2/21

はじめに

経済における波及
メカニズム

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

今回やったこと

データの拡張

実験計画法

実験計画法の手法

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.

本研究の背景 2

3/21

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に絡み合って影響を与えるかを詳細に分析できない。

研究の意義

専門的な知識がなくても為替リスクヘッジ戦略を最適化することができるようになる。

アプローチ

経済波及メカニズムに関する時系列データを分析

→ どの変数がどのように絡み合って為替に影響を与えてているのかを数理モデルで表す。

数理モデルの例

$$\text{USD/JPY} = 2.0 \cdot (\text{Nikkei225}) + 2.6 \cdot (\text{OIL}) - 1.0 \cdot (\text{VIX})$$

はじめに

経済における波及
メカニズム

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

今回やったこと

データの拡張

実験計画法

実験計画法の手法

おわりに

経済における波及メカニズム

4/21

はじめに

経済における波及
メカニズム

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

今回やったこと

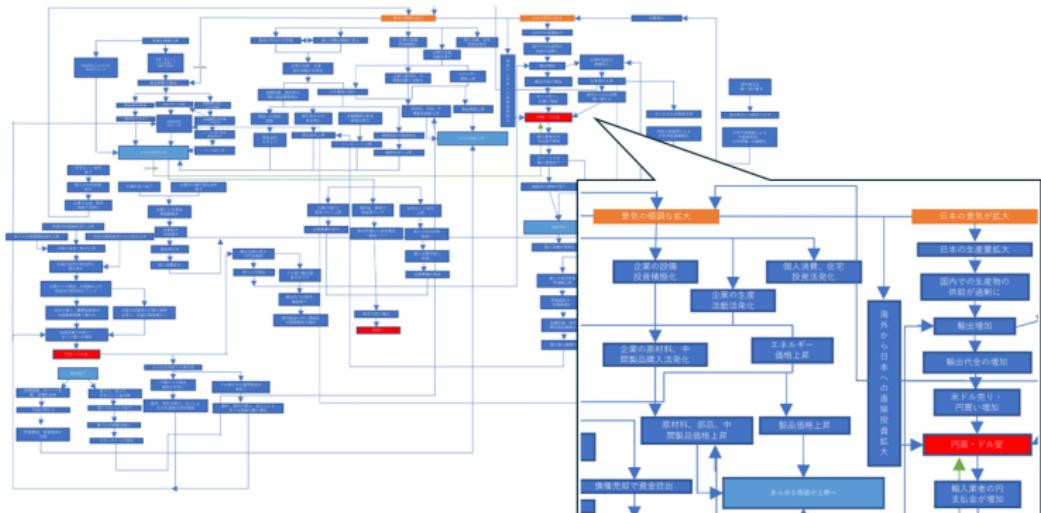
データの拡張

実験計画法

実験計画法の手法

おわりに

経済における様々な要因間の関係を表す経済の波及メカニズムがある。



はじめに

経済における波及
メカニズム

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

今回やったこと

データの拡張

実験計画法

実験計画法の手法

おわりに

公開されているデータ

経済波及メカニズムを参考に用いるデータを調査したところ、約 50 の変数が収集可能であった。

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

経済における波及
メカニズム

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

今回やったこと

データの拡張

実験計画法

実験計画法の手法

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

経済における波及
メカニズム

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

今回やったこと

データの拡張

実験計画法

実験計画法の手法

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

経済における波及
メカニズム

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

今回やったこと

データの拡張

実験計画法

実験計画法の手法

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) $y = x + 2 \cdot \sin(z) \rightarrow + x * 2 \sin z$ 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 によるシンボリック回帰

はじめに

経済における波及
メカニズム

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

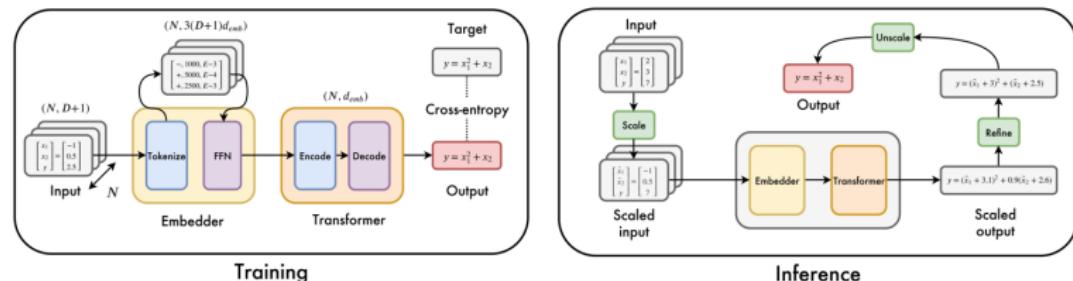
今回やったこと

データの拡張

実験計画法

実験計画法の手法

おわりに



Embedder

入力データをトークン化する.

埋め込みルックアップテーブルを使ってトークンをベクトルに変換する
(512 次元).

ベクトルを FFN に入力し, 短いベクトルに圧縮する.

この処理をすべてのデータ点に対して行う.

これを Transformer 本体に渡す.

Transformer のメカニズム

Transformer はエンコーダーとデコーダーから成り、Attention メカニズムが核を担っている。

■ 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。

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

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

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

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

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

経済における波及
メカニズム

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

今回やったこと

データの拡張

実験計画法

実験計画法の手法

おわりに

はじめに

経済における波及
メカニズム

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

今回やったこと

データの拡張

実験計画法

実験計画法の手法

おわりに

■ Feed-Forward Network

ベクトルを圧縮する際やベクトルを拡張する際に使われる。

Transformer モデルでは両方の使われ方がされている。

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (5)$$

ここで x は前のステップからの入力, W_1, W_2, b_1, b_2 はモデルが学習する重みとバイアスを意味する。

エンコーダー

入力されたデータ点を読み込み, それらがどんな関数の特徴を持っているかを分析する。

はじめに

経済における波及
メカニズム

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

今回やったこと

データの拡張

実験計画法

実験計画法の手法

おわりに

エンコーダー

Embedder によってベクトル化された N 個のデータ点を受け取る.

関連性の計算 (self-attention) を行い, 各データ点がほかのすべてのデータ点とどの程度関連しているかを計算する.

FFN によって関連性を加味した情報をさらに深く処理する.

関連性の計算と FNN による処理を 4 層繰り返す.

入力データ全体の特徴を要約した情報をデコーダーに渡す.

エンコーダーの self-attention

入力された全データ点の, 相互の関連性の強さを計算する.

自分自身を含むすべての入力データ点を参照する.

これによって各データ点のベクトルにデータセット全体における他の点との関係性を埋め込む.

デコーダー

エンコーダーからの情報と数式の一部を受け取る (初回は開始トークンのみを受け取る).

自己参照 (初回は開始トークンのみ, 2回目以降はトークン列を受け取る)

エンコーダーからの要約情報に注目し, 作る関数を理解する.

上記2角情報をもとに, 数式の次に来るべきトークンを予測する.

自己参照から予測までの流れを終了トークンが出力されるまで1トークンずつ繰り返す.

完成した数式を出力する.

デコーダーの自己参照

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} + M\right)V \quad (6)$$

注目しているトークンとほかの全トークンとの関連度を計算する.

マスクを利用して未来のトークンに関する部分を $-\infty$, それ以外の部分を 0 にする.

これに softmax 関数を適用することでモデルが過去と現在のトークンのみを参照して次のトークンを予測するようにする.

はじめに

経済における波及
メカニズム

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

今回やったこと

データの拡張

実験計画法

実験計画法の手法

おわりに

はじめに

経済における波及
メカニズム

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

今回やったこと

データの拡張

実験計画法

実験計画法の手法

おわりに

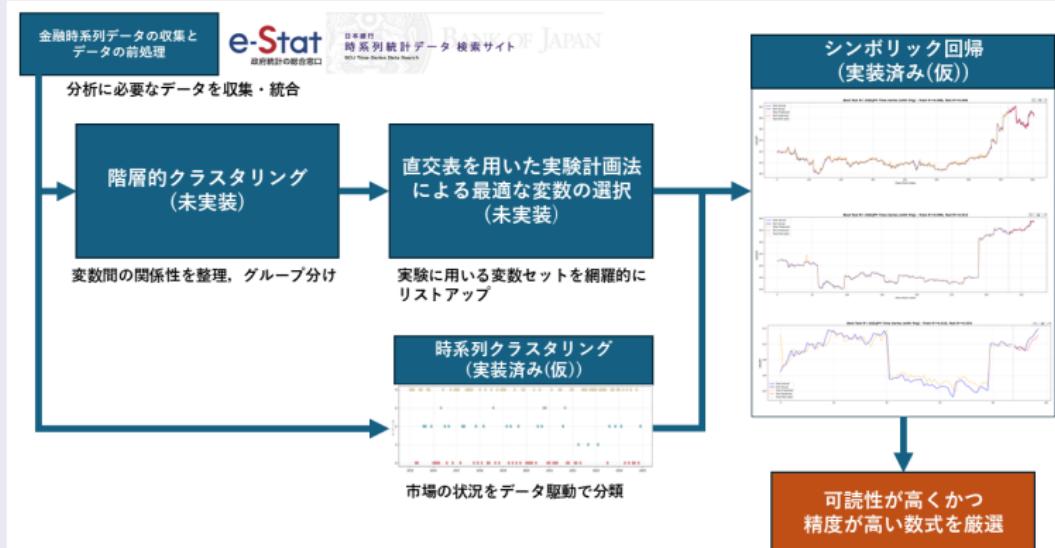
クロスエントロピーによる数式の評価

$$H(p, q) = - \sum_i p(x_i) \log q(x_i) \quad (7)$$

ここで $p(x_i)$ は正解の確率分布, $q(x_i)$ はモデルが予測した確率分布を意味する。

これを計算することで生成した数式の評価を行い, これを高めていくよう
に学習していってモデルを作る.

システム概要



今回やったこと

16/21

はじめに

経済における波及
メカニズム

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

今回やったこと

データの拡張

今回やったこと

データの拡張

空間充填計画を用いた実験計画の作成

一連のシステムの作成

実験計画法

実験計画法の手法

おわりに

データの拡張

17/21

課題

時系列クラスタリングを行うと、レジーム間でサンプル数に差があった
→ サンプル数が少ないレジームでは精度、可読性ともに低かった

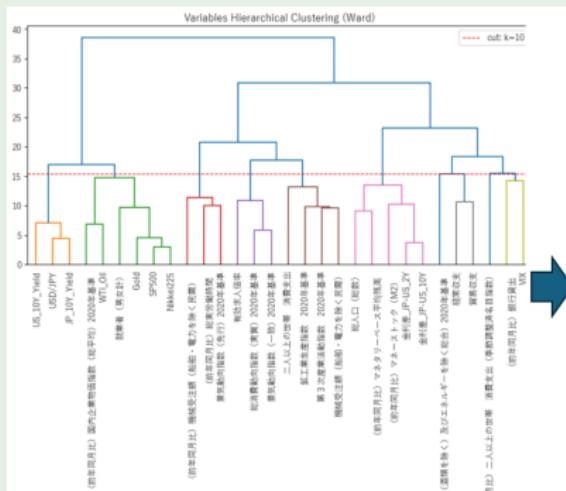


サンプル数の拡張

純粋にサンプル数を増やすためにデータの期間を拡張した。
2015-2025 だったものを 2007-2025 に
データ数は一旦 24 個に
必要であればより多くのデータを取得することも可能

実験計画法

実験計画法



	Cluster_1	Cluster_2	Cluster_3	Cluster_4	Cluster_5	Cluster_6	Cluster_7
0	Nikkesi225_1ag1	貿易収支 _lag1	(前年同月比) マネーストック (M2) _lag1	(前月比) 二人以上 の上の賃金 消費支 出 (季節調整済名 目指標)_lag1	VIXlag1	(前年同月比) マネーリバース 平均残高_lag1	(前年同月比) 機械・設備投資 (施 設・労働力を基 づき)_lag1
1	USD/JPYlag1	貿易収支 _lag1	(前年同月比) マネーストック (M2) _lag1	(前月比) 二人以上 の上の賃金 消費支 出 (季節調整済名 目指標)_lag1	VIXlag1	金利差_JP- US_2Ylag1	(前年同月比) 機械・設備投資 (施 設・労働力を基 づき)_lag1
2	S&P500_lag1	経常収支 _lag1	(前年同月比) マネーストック (M2) _lag1	(前月比) 二人以 上の賃金 消費支 出 (季節調整済名 目指標)_lag1	VIXlag1	総人口 (施設) _lag1	景気動向指数 (先行) 2020年 基準_lag1
3	Nikkesi225_1ag1	貿易収支 _lag1	(前年同月比) マネーストック (M2) _lag1	(前月比) 二人以 上の賃金 消費支 出 (季節調整済名 目指標)_lag1	VIXlag1	(季節調整済) GDP 工業生産指 数 2020年基準_lag1	(前年同月比) 機械・設備投資 (施 設・労働力を基 づき)_lag1
4	Gold_lag1	経常収支 _lag1	(前年同月比) 銀行貸出_lag1	(前月比) 二人以 上の賃金 消費支 出 (季節調整済名 目指標)_lag1	VIXlag1	総人口 (施設) _lag1	景気動向指数 (先行) 2020年 基準_lag1
5	就業者(両面 計)_lag1	貿易収支 _lag1	(前年同月比) マネーストック (M2) _lag1	(前月比) 二人以 上の賃金 消費支 出 (季節調整済名 目指標)_lag1	VIXlag1	総人口 (施設) _lag1	景気動向指数 (先行) 2020年 基準_lag1
6	Nikkesi225_1ag1	貿易収支 _lag1	(前年同月比) 銀行貸出_lag1	(前月比) 二人以 上の賃金 消費支 出 (季節調整済名 目指標)_lag1	VIXlag1	金利差_JP- US_10Ylag1	景気動向指数 (先行) 2020年 基準_lag1
7	Nikkesi225_1ag1	経常収支 _lag1	(前年同月比) 銀行貸出_lag1	(前月比) 二人以 上の賃金 消費支 出 (季節調整済名 目指標)_lag1	VIXlag1	総人口 (施設) _lag1	景気動向指数 (先行) 2020年 基準_lag1

実験計画法の種類

19/21

はじめに

経済における波及
メカニズム

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

今回やったこと

データの拡張

実験計画法

実験計画法の手法

おわりに

採用するアプローチ

今回の対象データはクラスタリングによって分割されたデータ
→ それぞれのクラスタにおいて異なる数のデータ、またレジームによって
クラスタリング結果も変わってくる。
そのため、空間充填計画による実験計画の作成を行う。

空間充填計画

組み合わせ空間全体を均一にサンプリングする手法
クラスタに属する変数間に偏りがあってもそれを考慮した実験計画を作成
できる。
勉強中

一連のシステムの実装 1

20/21

一連のシステムの実装 1

はじめに

経済における波及
メカニズム

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

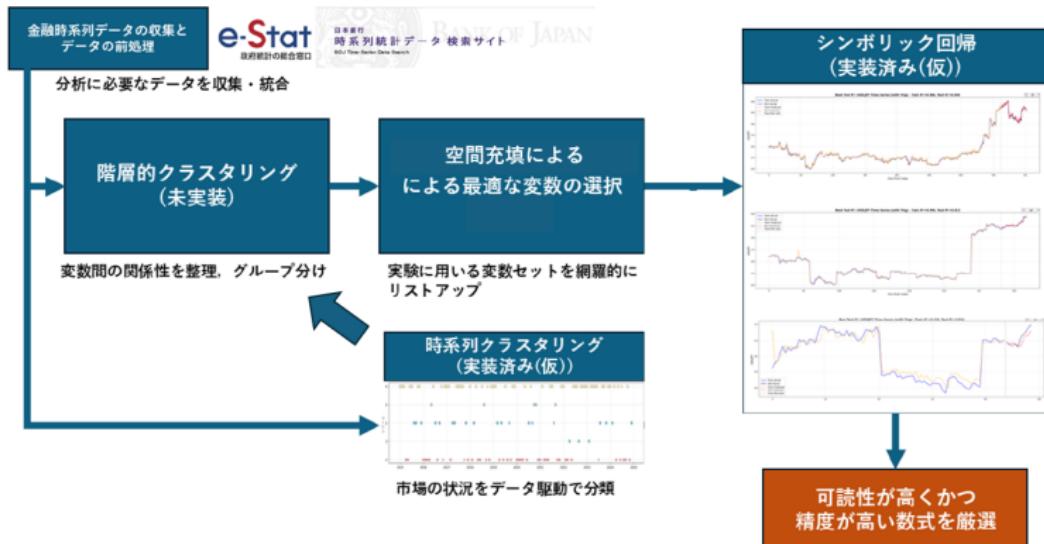
今回やったこと

データの拡張

実験計画法

実験計画法の手法

おわりに



はじめに

経済における波及
メカニズム

統計データの特徴
と研究の概要

End to end
symbolic
regression with
transformers

今回やったこと

データの拡張

実験計画法

実験計画法の手法

おわりに

まとめ

今回はシステムの一連の流れを実装した。

- 実験計画法の検討と実装
- データの拡張

今後の展望

- 時間が足りなくて実験まではできなかつたので、早速今日からシステムを動かす
- 可読性の定義と選ばれるの厳選について情報収集
- 実験で新規性を検証する。
- 論文を書きつつ、手法について勉強、プログラムの確認。