

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

階層的クラスタリ
ングについて

階層的クラスタリ
ングについて

数値実験並びに
考察

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

階層的クラスター
リングについて

階層的クラスター
リングについて

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

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

階層的クラスタリ
ングについて

階層的クラスタリ
ングについて

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に絡み合ってい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。

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

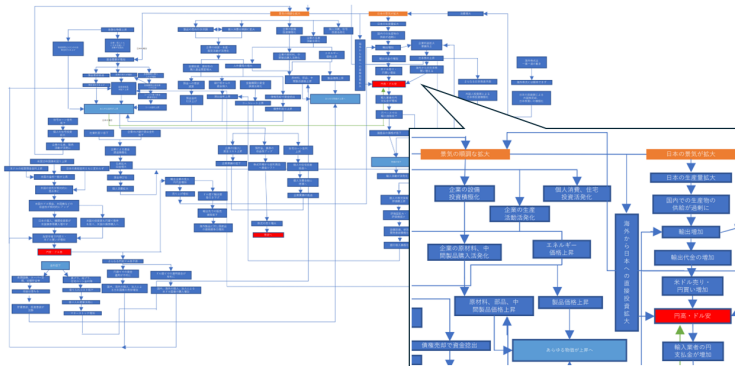
→ どの変数がどのように絡み合っているかを数理モデルで表す。

数理モデルの例

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

経済における波及メカニズム

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

階層的クラスタリ
ングについて

階層的クラスタリ
ングについて

数値実験並びに
考察

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

階層的クラスタリ
ングについて

階層的クラスタリ
ングについて

数値実験並びに
考察

終わりに

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

階層的クラスタリ
ングについて

階層的クラスタリ
ングについて

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

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

階層的クラスタリ
ングについて

階層的クラスタリ
ングについて

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:

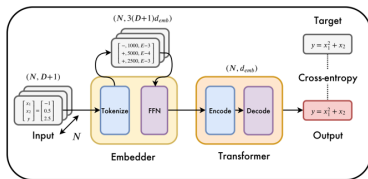
数式は、演算子、定数、変数といった要素が並んだシーケンスとして表現する.

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.

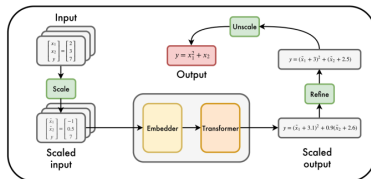
Transformer によるシンボリック回帰 (前準備)

9/23

Transformer によるシンボリック回帰



Training



Inference

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。

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

階層的クラスタリ
ングについて

階層的クラスタリ
ングについて

数値実験並びに
考察

おわりに

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。

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

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

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

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

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

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

階層的クラスタリ
ングについて

階層的クラスタリ
ングについて

数値実験並びに
考察

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

階層的クラスタリ
ングについて

階層的クラスタリ
ングについて

数値実験並びに
考察

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

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

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

階層的クラスタリ
ングについて

階層的クラスタリ
ングについて

数値実験並びに
考察

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たモデルを作る.

システム概要



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

階層的クラスタリングについて

階層的クラスタリングについて

数値実験並びに考察

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

階層的クラスタリ
ングについて

階層的クラスタリ
ングについて

数値実験並びに
考察

終わりに

今回やったこと

階層的クラスタリングの実装

直交表を用いた実験計画法の実装

一連のシステムについての実験

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

階層的クラスタリ
ングについて

階層的クラスタリ
ングについて

数値実験並びに
考察

おわりに

階層的クラスタリングによる変数選択手法

経済分析をするうえで非常に多くの変数を考慮する必要がある。

しかし、現在用いているシンボリック回帰「SymbolicRegression with Transformers」では 10 個の変数までしか取り扱えない。

→ シンボリック回帰に用いる変数で重要なものを 10 個選ぶ必要がある。

階層的クラスタリング

階層的クラスタリングでは相関の高いデータ同士を同一クラスタに分類することができる。

これによって多重共線性の原因をデータに基づいて特定することができる。

階層的クラスタリング

2つの時系列データ $X = (x_1, x_2, \dots, x_T)$ と $Y = (y_1, y_2, \dots, y_T)$ があるとき、以下の式で変数間の類似度 ρ を計算する。

$$\rho(X, Y) = \frac{\sum_{t=1}^T (x_t - \mu_X)(\mu_t - \mu_Y)}{\sqrt{\sum_{t=1}^T (x_t - \mu_X)^2} \sqrt{\sum_{t=1}^T (y_t - \mu_Y)^2}} \quad (8)$$

ここで、 T は時点の総数、 μ_X と μ_Y はそれぞれの時系列データの平均値を意味する。

次に以下のように絶対値を使って、類似度を距離へ変換させる。

$$d(X, Y) = 1 - |\rho(X, Y)| \quad (9)$$

次にウォード法を用いてクラスターの併合を行う。

$$\Delta ESS(i, j) = \frac{n_i n_j}{n_i + n_j} \|c_i - c_j\|^2 \quad (10)$$

ここで n_i, n_j はそれぞれのクラスターに含まれる変数の数、 c_i, c_j はそれぞれのクラスターの重心である。

使用したデータ

yfinance のデータ 10 個, e-stats から取得したデータ 20 個の計 30 個のデータを用いた.

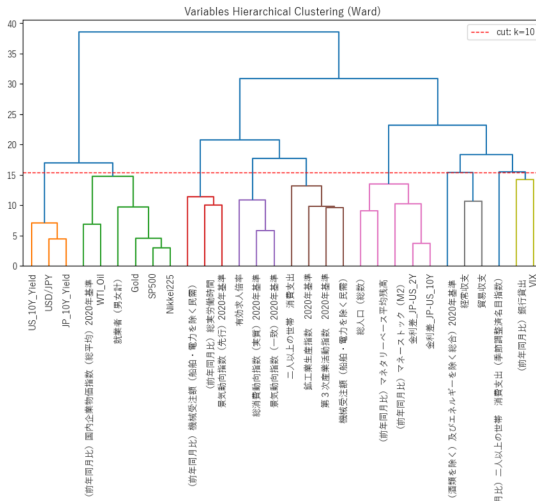
yfinance から取得したデータは日足だが e-stat から取得したデータはすべて月足データ.

月足を日足と混同して実験を行うと日足と月足で別のクラスタに分けられてしまい, 本来意図しない結果になってしまう.

→ 日足データを平滑化し, 月足データに直して実験を行う.

Table 2: 数値実験に用いたデータ

日足データ一覧	
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



実験計画法の作成

階層的クラスタリングの結果をもとに実験計画法を作成するシステムを作成した。
シンボリック回帰では 10 個の変数に絞り込む必要がある。

→10 個のクラスタから一つずつ変数を取り出し、実験計画法を作成する。

	Cluster_1	Cluster_2	Cluster_3	Cluster_4	Cluster_5	Cluster_6	Cluster_7	Cluster_8	Cluster_9	Cluster_10
0	JP_10Y_Yield	Nikkei225	(前年同月比) 総実労働時間	景気動向指数 (一致) 2020年基準	鉱工業生産指数 2020年基準	総人口 (総数)	(前月比) 消費者物価指数 (食料 (酒類を除く) 及びエネルギーを除く組合) 2020年基準		(前年同月比) 銀行貸出	(前月比) 二人以上の世帯消費支出 (季節調整済み指標)
1	USD/JPY	SP500	(前年同月比) 総実労働時間	景気動向指数 (一致) 2020年基準	二人以上の世帯消費支出	総人口 (総数)	(前月比) 消費者物価指数 (食料 (酒類を除く) 及びエネルギーを除く組合) 2020年基準		VIX	(前月比) 二人以上の世帯消費支出 (季節調整済み指標)
2	JP_10Y_Yield	SP500	(前年同月比) 総実労働時間	有効求人倍率	二人以上の世帯消費支出	総人口 (総数)	(前月比) 消費者物価指数 (食料 (酒類を除く) 及びエネルギーを除く組合) 2020年基準		VIX	(前月比) 二人以上の世帯消費支出 (季節調整済み指標)
3	JP_10Y_Yield	SP500	景気動向指数 (先行) 2020年基準	有効求人倍率	二人以上の世帯消費支出	金利差 JP-US_2Y	(前月比) 消費者物価指数 (食料 (酒類を除く) 及びエネルギーを除く組合) 2020年基準		(前年同月比) 銀行貸出	(前月比) 二人以上の世帯消費支出 (季節調整済み指標)
4	JP_10Y_Yield	Nikkei225	景気動向指数 (先行) 2020年基準	有効求人倍率	鉱工業生産指数 2020年基準	総人口 (総数)	(前月比) 消費者物価指数 (食料 (酒類を除く) 及びエネルギーを除く組合) 2020年基準		VIX	(前月比) 二人以上の世帯消費支出 (季節調整済み指標)
5	USD/JPY	Nikkei225	景気動向指数 (先行) 2020年基準	景気動向指数 (一致) 2020年基準	二人以上の世帯消費支出	総人口 (総数)	(前月比) 消費者物価指数 (食料 (酒類を除く) 及びエネルギーを除く組合) 2020年基準		(前年同月比) 銀行貸出	(前月比) 二人以上の世帯消費支出 (季節調整済み指標)
6	USD/JPY	Nikkei225	(前年同月比) 総実労働時間	有効求人倍率	二人以上の世帯消費支出	金利差 JP-US_2Y	(前月比) 消費者物価指数 (食料 (酒類を除く) 及びエネルギーを除く組合) 2020年基準		(前年同月比) 銀行貸出	(前月比) 二人以上の世帯消費支出 (季節調整済み指標)
7	USD/JPY	SP500	景気動向指数 (先行) 2020年基準	有効求人倍率	鉱工業生産指数 2020年基準	総人口 (総数)	(前月比) 消費者物価指数 (食料 (酒類を除く) 及びエネルギーを除く組合) 2020年基準		(前年同月比) 銀行貸出	(前月比) 二人以上の世帯消費支出 (季節調整済み指標)
8	JP_10Y_Yield	Nikkei225	景気動向指数 (先行) 2020年基準	景気動向指数 (一致) 2020年基準	二人以上の世帯消費支出	金利差 JP-US_2Y	(前月比) 消費者物価指数 (食料 (酒類を除く) 及びエネルギーを除く組合) 2020年基準		VIX	(前月比) 二人以上の世帯消費支出 (季節調整済み指標)
9	USD/JPY	Nikkei225	(前年同月比) 総実労働時間	有効求人倍率	鉱工業生産指数 2020年基準	金利差 JP-US_2Y	(前月比) 消費者物価指数 (食料 (酒類を除く) 及びエネルギーを除く組合) 2020年基準		VIX	(前月比) 二人以上の世帯消費支出 (季節調整済み指標)
10	USD/JPY	SP500	景気動向指数 (先行) 2020年基準	景気動向指数 (一致) 2020年基準	鉱工業生産指数 2020年基準	金利差 JP-US_2Y	(前月比) 消費者物価指数 (食料 (酒類を除く) 及びエネルギーを除く組合) 2020年基準		VIX	(前月比) 二人以上の世帯消費支出 (季節調整済み指標)
11	JP_10Y_Yield	SP500	(前年同月比) 総実労働時間	景気動向指数 (一致) 2020年基準	鉱工業生産指数 2020年基準	金利差 JP-US_2Y	(前月比) 消費者物価指数 (食料 (酒類を除く) 及びエネルギーを除く組合) 2020年基準		(前年同月比) 銀行貸出	(前月比) 二人以上の世帯消費支出 (季節調整済み指標)

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

階層的クラスタリングについて

階層的クラスタリングについて

数値実験並びに考察

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

階層的クラスタリ
ングについて

階層的クラスタリ
ングについて

数値実験並びに
考察

終わりに

考察

今回は実験を行う上で都合がよいように 10 個のクラスタに分けて実験した。
適切なクラスタ数を設定し、実験を行うことでより違った結果が見えてくると考えら
れる。

日足と月足はある程度分けられてはいるが、それでも少し偏っているように感じら
れる。

同一クラスタに分けられているデータ

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

階層的クラスタリ
ングについて

階層的クラスタリ
ングについて

数値実験並びに
考察

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.

- 階層的クラスタリングによる説明変数の絞り込み
- 実験計画を作成するプログラムの作成

今後の展望

- より深い実験を行い, 新規性を発見する.
- 新規性が発見できたなら修士論文に執筆.
- 新規性が見つからなければ実装する.