

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

データについて

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

データについて

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。

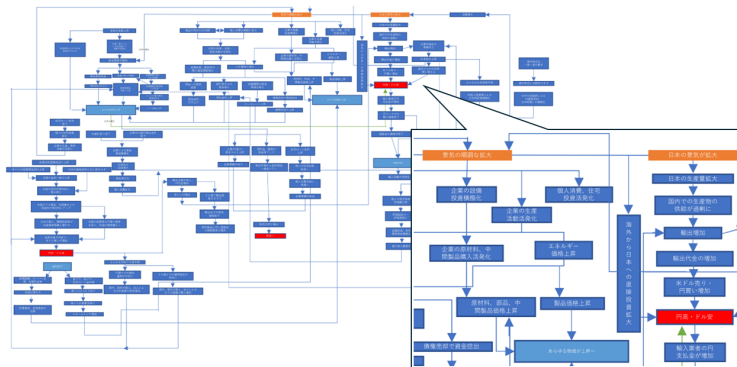
数理モデルの例

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

経済における波及メカニズム

4/20

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

データについて

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

データについて

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

データについて

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

データについて

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

データについて

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:

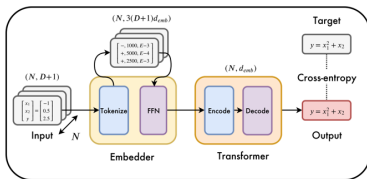
数式は、演算子、定数、変数といった要素が並んだシーケンスとして表現する.

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.

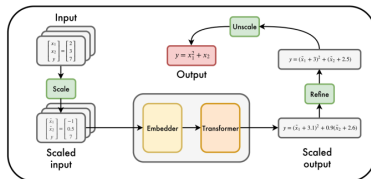
Transformer によるシンボリック回帰 (前準備)

9/20

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

データについて

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

データについて

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

データについて

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。

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

データについて

おわりに

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

データについて

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.

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

データについて

おわりに

今回やったこと

修論の方針決定

データ収集の実装.

データ選択に使うクラスタリング手法の検討.

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

データについて

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

データについて

おわりに

モデルベースの時系列クラスタリング

経済波及メカニズム中にあるデータを可能な限り多く取得する必要がある。
経済波及メカニズムでは労働力や輸出入金額など財政のデータが非常に多い。

マネーストックや金利などのデータは日本銀行時系列検索サイトから取得するが、それ以外のデータは e-stat から取得する必要がある。

e-stat について

e-stat では財政データなどを多く公開しており、中でも統計 Dashboard では時系列データを多く公開している。

今回はこのデータ取得を行った。

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

データについて

おわりに

統計 Dashboard

API でデータを公開しており、

e-stat について

e-stat では財政データなどを多く公開しており、中でも統計 Dashboard では時系列データを多く公開している。

今回はこれのデータ取得を行った。

まずは API で取得したいデータの系列コードを調べ、時間足や開始日などを指定。これによってデータを呼び出し、Google スプレッドシートに記録するようプログラムを作成した。

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

データについて

おわりに

階層的クラスタリング

シンボリック回帰に用いる 10 個のデータを絞り込むために行うクラスタリングとして階層的クラスタリングを考える。

階層的クラスタリングはクラスタリング手法のひとつで最も似ているデータから順に一つずつペアにしてクラスタを作る手法。

これを行ううえでもさらに様々な手法が存在するが、この研究ではワード法を使ってクラスタリングを行う。

ワード法ではクラスタ内のばらつきが最も小さくなるようにクラスタリングを行うため、相関が高い変数同士が優先的にまとめられる。

時系列の考慮について

一般的なワード法では時系列を考慮できていないという欠点があるため、今回の研究でそのまま使うことができない。

そのため、動的伸縮手法 (DTW) を用いて時系列に対応させる。

これについてはすでに研究されているため、そのアルゴリズムを使っていく。

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

データについて

おわりに

まとめ

今回は経済波及メカニズムにある変数で取得できそうなものをすべて取得するプログラムを作成した。

また、変数の絞りこみに使う手法を検討した。

今後の展望

- 時系列データを取得するプログラムの実装はできたので、次は階層的クラスタリングを実際に実装する。
- 実装ができたらずべて一貫させたプログラムを作成し、精度や可読性を評価する。
- 精度や可読性に新規性が見られたら論文を執筆する。