

1. はじめに

修士研究のテーマについて z

Adjusting Input Coefficients and Visualising Spillover Paths
in an Input-Out Analysis from Financial and Economic Data

蒲田 涼馬 (Ryoma Gamada)
u2455007@st.pu-toyama.ac.jp

富山県立大学 工学研究科
電子・情報工学専攻 情報基盤工学講座

May 27, 2025

1. 金融経済データからの産業連関分析における 投入係数の調整と波及経路の可視化

2/12

背景

近年、金融工学は計算機性能の向上やデータサイエンス手法の進化、公的機関によるオープンデータの提供に伴い、飛躍的な発展を遂げている、将来予測などの分野は成長しているにも関わらず、多様な要素の相互関係や経済の動向を可視化するような研究は多くない。



将来予測: 資源配分の最適化やリスク管理の向上



発展

影響分析: 政策分析や経営における戦略分析などに使われる



少ない

Figure 1: 研究の背景

1. はじめに

目的

AI, あるいは別の手法を用いて過去の金融データから自動的に法則を発見するシステムを作成する.

なるべく金融に特化したものを考え, 最終的には他の手法との比較を行う.
発見された数式の可読性を評価と金融市場の理解に貢献できる知見を得る.

1. はじめに

シンボリック回帰 (Symbolic Regression)

シンボリック回帰はデータセットに適合する数学的表現を自動的に探索する機械学習の手法.

従来の回帰分析がモデル構造を事前に仮定するのに対し, 記号的回帰はモデル構造自体を発見する.

PySR

PySR では損失関数を以下のように定義している.

$$\mathcal{L}(E) = \mathcal{L}_{pred}(E) + \lambda \cdot C(E) \quad (1)$$

ここで E は候補となる数式, $\mathcal{L}_{pred}(E)$ は予測誤差, $C(E)$ は式の複雑度, λ は複雑性に対するペナルティ係数を意味する.

これによって可読性, 精度の両方が高い式を生存させる.

一般的な進化アルゴリズム

1. 準備: 多くの数式の集まりを用意する. 個体の良し悪しを測る基準と個体を変化させるルールを設定する.
2. トーナメントによる個体の厳選: 集団の中からランダムに数個の個体 (サブセット) を選ぶ. サブセット内で適合度を評価し, 競争させる.
3. 新しい個体の生成: トーナメントで選ばれた個体をコピーし, そのコピーにランダムに突然変異を加えて少し変化させる.
4. 集団の更新: 新しく生成した個体とサブセット内で最も劣っていた個体を入れ替える.

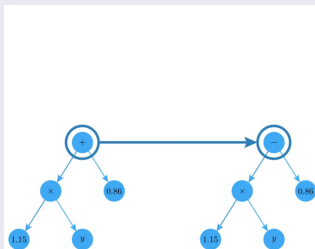


図3. 突然変異動作

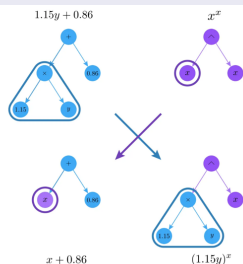


図4. 交叉動作

PySR

1. 準備: ランダムな「数式」を多く生成. PySR ではこの集団をさらに islands に分けてそれぞれで進化を進める.
2. 数式の評価: 損失関数の計算を行い各数式がどれだけの優劣なのかを評価する.
3. 数式の選択: 点数の高い数式が次の世代の親として選ばれやすくなる.
4. 新しい数式の作成: 選ばれた親の数式を元に新たな数式を生成する. ここで交叉, 突然変異, 定数最適化を行う.
5. 数式集団の入れ替え: 新しくできた数式と元の集団の劣っている数式を入れ替える.
6. islands 間の交流: 定期的に各 islands から優秀な数式が他の islands へ移動する.
7. 繰り返しと最良の数式の選択

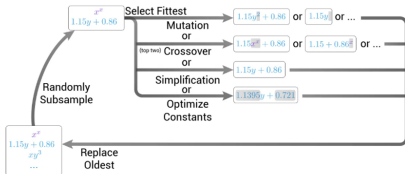


図3. PySRの内部ループ

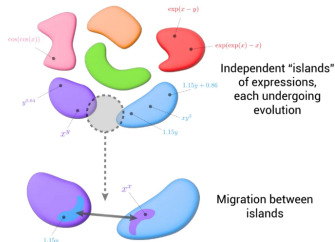


図4. PySRの外部ループ

1. はじめに

実装について

以前、長瀬さんが提案した手法と PySR との比較をとりあえず目視で行う。
長瀬さんが提案した手法では潜在プロファイル分析を行いパーセプトロンを用いた数
法則発見を行っている。

対象としたデータについて

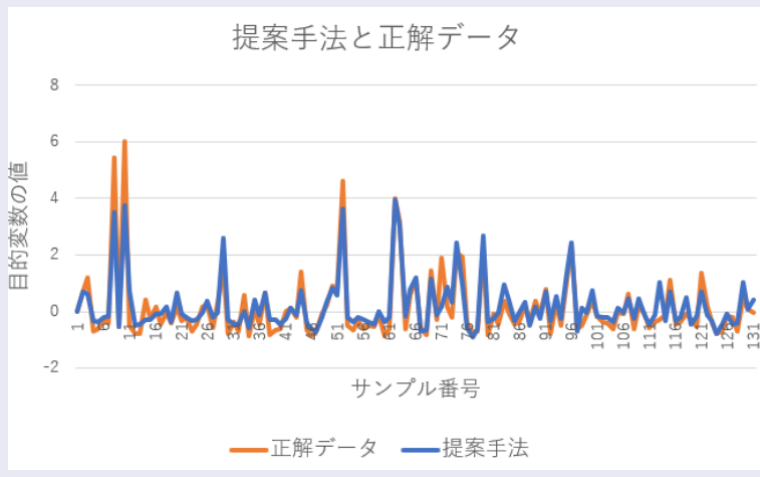
一人当たりの固定資産税，一人あたりの地方税，一人あたりの法人住民税，経営耕地
面積，製造品出荷額，年間商品販売額，総人口，住宅用地平均取引価格，商業用地平
均取引価格，林地平均取引価格の計 11 種類の量的データ。

実験概要

データ数は 650 であり，ここからランダムに 130 個の変数を抜き出してそれをテス
トデータ，それ以外のデータをトレーニングデータとして扱う。

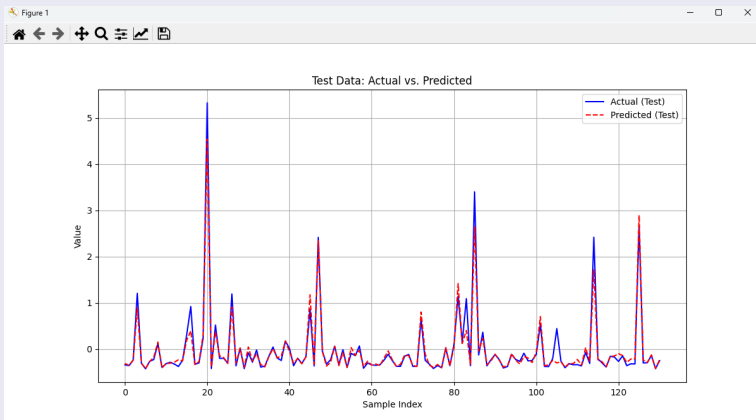
結果 1

まず長瀬さんの提案手法の結果を以下に示す。



結果 2

PySR を同じデータに適用させた結果を以下に示す。



1. はじめに

結果と考察

どちらもある程度テストデータでもフィットできていると感じるが, 実際どちらが優れているのかがまだわからない.

長瀬さんのプログラムを自分自身で動かし全く同じ条件で MSE(二乗平均誤差) を比較する必要がある.

PySR の方では一部ピークを取れていないところがある.

→取りこぼしが無いように (局所にも強い)PySR を組み替えていくことができれば精度が向上し新規性になるかもしれない.

1. はじめに

CNN を用いた予測について

畳み込みニューラルネットワークを用いた株価の予測が近年のトレンド。
今回は基本的な CNN による日経平均株価予測システムを実装してみた。
データについては予測する変数を日経平均株価の終値, 説明変数は日経平均株価の始値, 高値, 低値, 過去の終値, 出来高, ドル円の終値, GSPC の終値を用いる。
データは 2015 年 1 月 1 日から 2025 年 1 月 1 日までの 10 年間のデータを用いて, その後半 2 割をテストデータとした。
データの予測には過去 20 日のデータを用い, エポック数は 40, バッチサイズは 32 で実験を行った。

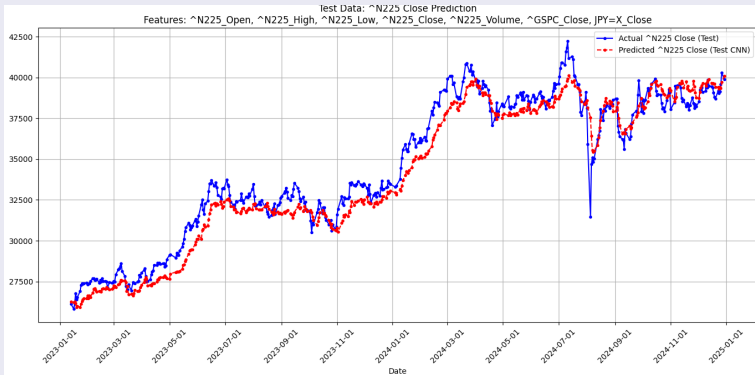
モデルについて

2 つの畳み込みブロック (Conv1D + MaxPooling1D + Dropout) で時系列データから特徴を抽出し, 平坦化し, 最終的に 1 つの値を予測する構造でモデルを構築した。

結果について

1. はじめに

予測結果は以下のようになり、ある程度概形は捉えられている程度の結果になった。しかし予測値が正解データより少し右にずれているように感じられる。



1. はじめに

まとめ

PySR と長瀬さんの提案手法との比較を行った。
比較は十分ではなくはっきりとした違いは今のところ見つけられなかった。
CNN による株価の予測もしてみたが、基本的なモデルの実装だったため精度はそんなに高くなさそう。

今後やること

PySR と長瀬さんの手法を全く同じ条件で数値的に比較する実験を行う。
CNN についての最新の取り組みについて調査し、実際に実装し新規性について考える。