

# 経済金融時系列における歪みの検出

## Detection of distortion in economic and financial time series

土屋 太助<sup>1</sup>

Taisuke Tsuchiya<sup>1</sup>

**Abstract:** The market is greatly disrupted by economic news and online media information such as twitter, causing distortions in the economic and financial time series of commodity futures markets. In this research, we applied the anomaly detection method by using of Dense Auto Encoder to economic and financial time series data to detect distortion of economic and financial time series such as commodity futures market.

### 1. はじめに

近年、金融データへの AI の応用が注目されており、経済金融時系列を Recurrent Neural Network (RNN) の入力として経済金融時系列の期先予測を行うといった研究[1]が行われているが、経済ニュースや twitter といったオンラインメディアの情報により、市場は大きく混乱し、各国の経済政策や地政学上の問題によってはこれまでの価格決定ロジックの常識が覆ることがある。また、オンラインメディアの情報を入力として自然言語処理により経済金融時系列の期先予測を行う研究[2]も行われているが、オンラインメディアの情報のすべてが常にターゲットの時系列に直接的に影響を与えるわけではなく、例えば、「NY ダウが暴落し金が高騰」といった、金融時系列からも得られる重複した二次情報も含まれている。金融時系列とオンラインメディアから抽出したスパースな特徴量を単純に結合しても、期先予測の精度向上は容易ではない。その時々相場感、経済環境、ニュースを鑑みて、専門家が特徴量を都度変更し、モデルを改修して運用することも可能だが、メンテナンス性、汎用性の観点から実用的ではない。加えて、今現在予測精度が高くても、そのモデルがいつまで、どういう状態であれば使い続けても問題ないのか、どういう状態になったら使用を停止すべきかの判断も難しい。データからこうした状況変化のシグナルを検出できれば、(部分的にでも) 期先予測モデルの運用開始を意思決定することができるのではないかと考えた。

本研究は、経済金融時系列から目的変数毎のレジームの転換点を検出し、レジーム毎に最適なモデル(モデルの使用停止を含む)を選択することを最終目的としている。今回はその中で、異常検出手法を

用いた経済金融時系列の歪みの検出を行った。

具体的には、センサデータの異常検知等[3]で用いられる自己符号化器(Auto Encoder)を用いて経済金融時系列の歪み(異常)を検出する。自己符号化器により次元圧縮し、入力と出力の差分である再構築誤差を金融時系列の歪みと定義し解析した。

次に、歪み情報を予測モデルの入力データと結合することで、予測精度改善効果が得られるか否かを検証した。

### 2. 自己符号化器による異常検知

オートエンコーダはニューラルネットワークの一種で、情報量を小さくした特徴表現を獲得する。オートエンコーダによる異常検知の仕組みは、エンコーダで入力(x)を、入力(x)より低次元な潜在変数(z)に次元圧縮し、デコーダで元の入力を復元する。学習は教師なし学習で、入力と出力の誤差が小さくなるように学習を行う。異常検知はこの入力と出力の誤差(再構築誤差)を用いる。(図1)

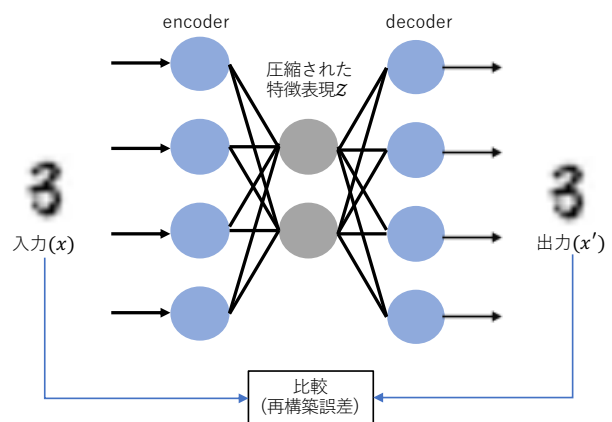


図1：自己符号化器による異常検知の仕組み

推論時、学習データに含まれない未知のデータ入力に対して再構築誤差が大きいときに異常として検出する仕組みである。センサデータや画像データの異常検知に用いられる手法だが、画像系の異常検知においては全結合層ではなく、畳み込み層

(Convolution 層) を、Deep なネットワーク構造とともに用いることが一般的である。Resnet や skip-connect を追加した U-net など復元能力の高いネットワーク構造は、未知のデータに対しても再構築能力が高いため、再構築残差が出にくく、結果として異常を検出できない。データおよび異常の特性を踏まえ、適度な復元力のネットワーク構造、ハイパーパラメータを選ぶ必要がある。

### 3. モデルへの入力特徴量

ニューヨーク商品取引所 (NYMEX) の原油先物価格をターゲットとし、金、プラチナ、パラジウムなどの先物価格に加えて日本、米国、中国、欧州の為替、米国、中国の主要な株価指数の計 38 種の日次データを特徴量として使用した。

### 4. ML アーキテクチャと具体的な実装

アーキテクチャ全体はデータ取得、前処理、予測の 3 つプロセスで構成しており、Dense Auto Encoder による次元圧縮および再構築残差の算出し、オートエンコーダのニューラルネットワークの構造は入力層、隠れ層 1 層、出力層という深層ではないネットワーク構造を用いた。歪みの検出という目的を鑑みて、適度な残差が発生するよう、全結合の隠れ層 1 層のシンプルなネットワーク構造を選択する。

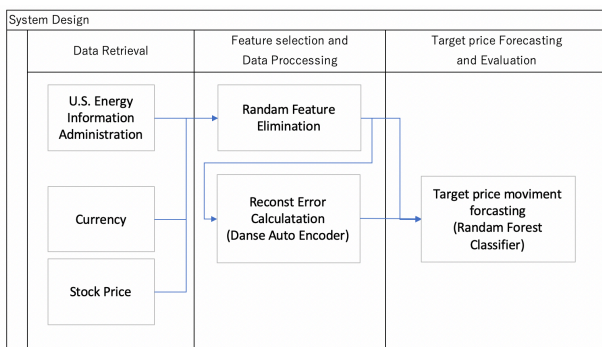


図 2 : ML アーキテクチャ概要

具体的なモデルとして下記を用いる。

1. Dense Auto Encoder: TensorFlow を用いて実装を行った。隠れ層 1 層の全結合型の構造で、活性

化関数は Tanh 関数を用いた。epoch 数は 200epoch とし、最適化アルゴリズムに Adam を用い、loss 関数は MSE を用いた。

2. Random Feature Elimination(RFE): scikit-learn の sklearn.feature\_selection を用いて実装を行った。
3. Random Forest (RF):パッケージ scikit-learn のクラス Random Forest Classifier を用いて実装を行った。

### 5. 結果と考察

まず、ターゲットである原油先物価格のテスト期間における再構築誤差の推移を図 3 にて示す。

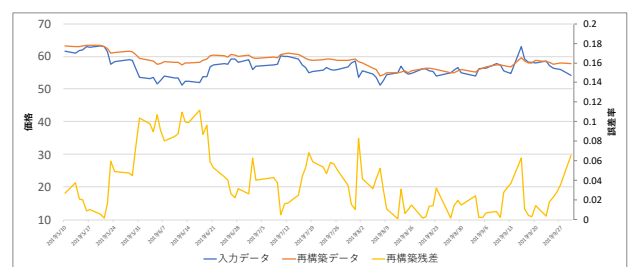


図 3 : テスト期間における再構築誤差の推移

Autoencoder の入力 38 次元に対して、8 次元まで次元圧縮しているため、テスト期間の時間が進むにつれて再構築誤差が発散する可能性もあると予想していたが、予想に反してテスト期間のターゲット先物価格の再構築誤差は一定以下に抑えられており、次元圧縮、特徴抽出は成功していると言える。

また、2019 年 5 月下旬から再構築誤差が 1 ヶ月程度大きくなっているが、これらの時期に、トランプ大統領による原油の増産圧力をかける発言がでており、実際に入力データに含まれていない米国在庫統計の情報を確認したところ供給量増 (価格下落要因) が確認できた。

次に、入力に在庫統計情報を加え、1 日先のターゲット価格の価格上昇、下落を Random Forest Classifier により 2 クラス分類問題として予測した。入力データに Dense Auto Encoder の次元圧縮後特徴量、再構築誤差を結合させ Random Forest Classifier モデルに入力した。評価結果(30 回評価の平均値)を図 4 に示す。入力+再構築誤差 (ターゲットの誤差のみ) を入力としたモデルがベースモデルよりもわずかに精度が高いという結果が出た。一方で全特徴量分の再構築誤差を入力に加えた場合、精度は向上しなかった。

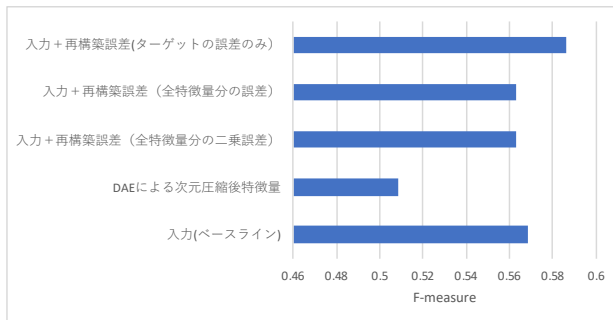


図4：DAE 特徴量を加えた際の F 値変化

## 6. まとめと今後の課題

本研究では、経済金融時系列データから歪みを検出し、レジーム毎にモデルを構築し、最適なモデルを自動選択することを最終目的として、今回はその中で、自己符号化器を用いた経済金融時系列からターゲット変数毎の歪みを抽出できることを確認した。今後の課題を下記に列挙する。

- ・自己符号化器を用いた **Out of Distribution** の検出
- ・再構築誤差を用いた解釈可能性の高いレジーム分解手法の構築
- ・レジーム毎の動的特徴選択とモデル構築

また、今回は、少ない入力（特徴次元数）で実験したが、**Autoencoder** による次元圧縮の効果を最大化するためには精度向上に寄与する意味のある特徴量の追加が必要である。

## 参考文献

- [1] Wei Bao, Jun Yue, Yulei Rao, A deep learning framework for financial time series using stacked autoencoders and long-short term memory(2017)
- [2] Xuerong Li, Wei Shang, Shouyang Wang : Text-based crude oil price forecasting: A deep learning approach(2018)
- [3] Chong Zhou, Randy C. Paffenroth : Anomaly Detection with Robust Deep Autoencoders, KDD '17 the 23rd ACM(2017)