同時進行する独立な時系列データからの Transformer による 予測モデルの構築

長谷川 雄大[†] 藤田 悟[†]

法政大学大学院 情報科学研究科[†]

1. はじめに

近年、ビッグデータの活用方法に注目が集まっている。活用方法の一例として、時系列データの予測がある。時系列データの予測は、活用の範囲が広く、例えば、過去の実績から将来の商品の需要を予測することができる。中でも、需要予測は欠品による販売機会の損失や、過剰な在庫による管理コストの削減につながるため、企業からの注目も高い。時系列データの予測において、誤差を少なく予測するためには、対象となる系列だけでなく、関連のある他の時系列データを活用することが重要である。そこで本研究では、時間軸方向の学習を行うモデルと、他時系列データとの関係を学習するモデルを組み合わせた予測モデルの構築を行った。

2. 関連研究

近年、時系列データの予測に Transformer を適応する手法に注目が集まっている [1]。Transformer は、元々、自然言語処理の分野で登場したモデルであり、再帰や畳み込みなどのコストのかかる計算を行わず、翻訳性能も高いことから注目を集めた [2]。[1] では、Transformer をベースとしたモデルでインフルエンザの感染者数を予測し、既存の時系列データに対する予測手法である ARIMA モデルや LSTM モデルよりも予測性能が向上したことを示した。

ニューラルネットワークを用いた予測において、予測の対象となる時系列データと相関のあるデータの特徴量として用いた場合、相関性を学習し、予測に貢献することが期待できる。しかし、Transformer を用いた予測において、特徴量に複数の時系列データを束ねて用いた場合、予測性能が落ちてしまうことが予備実験で判明した。

3. 提案手法

2章で述べたように、Transformer では、特徴量に複数の時系列データを束ねて用いた場合、

それがノイズとなり予測性能が落ちてしまうという問題点が存在する。そこで本研究では、まず、予測の対象となる時系列データごとに、そのデータのみを用いて時系列方向の学習を行い、その後、時系列方向の学習を終えたデータに対して、他の時系列データとの関係性を学習させることで、他の時系列データがノイズにならず、むしろ予測に貢献できるモデルを提案する。

3.1 問題設定

本稿では、 N 個の時系列データの T 時点までのデータを入力とし、その N 個の時系列データの $T+1$ 時点の値を予測する。

3.2 提案モデル

提案モデルはまず、Time Transformer で、時系列方向の学習を行う。その後、時系列方向の学習を行ったデータに対して、Space Transformer で時系列データ間の特徴量を学習する。そして、最後に、線形結合層を用いて値を出力する。Time Transformer と、Space Transformer については、以下で説明を行う。

(1) Time Transformer

Time Transformer では、入力されたデータ $X \in R^{N \times T \times D}$ の時系列方向に関して学習を行う。 N は時系列データの数、 T は時系列データの長さ、 D は特徴量の数を表している。時系列方向への学習は、それぞれの時系列データごとに独立して行われる。

まず、入力データに対して、[2]と同様に Positional Encoding により位置情報を付与する。

次に、Self-Attention によって、自身の過去のデータとの関連度を計算する。Time Transformer における Self-Attention では未来の情報を参照しないようにするためにマスクをかける。

その後は [2]と同様に、Feed Forward Network (FFN) により値を出力する。Time Transformer の出力 y_t は $y_t \in R^{N \times T \times D}$ となる。

(2) Space Transformer

Space Transformer では、それぞれの時刻に対して、その時刻における他データとの関係性を

Building a Forecasting Model with Transformer from Concurrent Independent Time Series Data

[†]Yuta Hasegawa, Satoru Fujita

[†]Graduate School of Information Sciences, Hosei University

学習させる。

最初の処理として、Time Transformer の出力 $y_t \in R^{N \times T \times D}$ の次元を入れ替える。入れ替えた後の値 $x_s \in R^{T \times N \times D}$ は、 $x_s = y_t^T$ となる。

次に、次元を入れ替えたデータに対して Self-Attention を行い、同時刻における他データとの関連度を計算する。Space Transformer の Self-Attention は、時刻ごとに独立して関連度を計算しているため、未来の情報を参照することはない。そのため、マスクは不要である。その後は Time Transformer と同様に FFN で値を出力する。

最後に、入れ替えていた次元を、再び転置することで元の次元に戻し出力する。すなわち、Space Transformer の出力 y_s は $y_s \in R^{N \times T \times D}$ となる。

4. 実験

4.1. 実験データ

アメリカあるフードデリバリー会社のデータを使用する^{*1}。このデータは、フードのオーダー数が週ごとに記録されている。フードは全部で45種類あり、オーダー数は145週間分ある。このデータの132週間分を学習に使用し、残りの12週間分のデータをテストデータに用いた。

4.2. 評価指標

モデルの評価には MAPE (Mean Absolute Percent Error) を用いる。MAPE は正解値と予測値の差を、正解値に対する割合で表すため、小さいほうが良いと言える。

4.3. 結果

提案モデルとその他のモデルで予測を行った結果を表1に示す。表中の ARIMA は ARIMA モデルによる予測結果であり、LSTM は LSTM を用いたモデルでの予測結果である。また、Transformer(単商品)は提案モデルの Time Transformer と、値を出力するための全結合層のみを用いたモデルであり、Transformer(多商品)は、Transformer(単商品)と同様のモデルで、入力データの特徴量次元に他商品の実績を使用したモデルである。提案手法(TTTS)は本稿で提案したモデルであり、今回の実験では、Time Transformer を3層重ねた後に、Space Transformer を使用した。ニューラルネットワークを用いたモデルは、重みとバイアスの初期値による学習結果のばらつきを抑えるために、それぞれ100回実験を行った。ARIMA モデルに関しては、初期値によるばらつきはないため、実験回数は1回である。表中の値は100回実験を行った時の MAPE の平均値と標準偏差である。実験の結果、提案手法による予測が、MAPE の値が最も小さくなった。Transformer(単商品)と

表1 100回の平均値と標準偏差

	MAPE	
	平均	標準偏差
ARIMA	0.358	-
LSTM	0.329	0.032
Transformer(単商品)	0.374	0.023
Transformer(多商品)	0.507	0.051
提案手法(TTTS)	0.308	0.016

Transformer(多商品)を比較すると、特徴量次元で他商品を用いた場合、予測誤差は約13%増加しており、予測性能が大きく低下している。一方で、提案した手法で他商品の情報を用いた場合、Transformer(単商品)と比較して、予測誤差は約6.6%減少した。また、提案手法は ARIMA や LSTM と比較しても予測性能が改善されており、ARIMA で約5.0%、LSTM では約2.1%予測誤差が減少した。

5. 考察

Transformer(単商品)とTransformer(多商品)の比較結果から、特徴量次元で他商品のデータを用いた場合、それがノイズとなってしまう、予測性能が低下したと考えられる。一方で、提案手法では、他商品の情報がノイズとならず、むしろ予測に貢献したことで、予測性能の向上につながったと考えられる。これは、Space Transformer で他の時系列データが予測の対象となる時系列データに対する影響度を学習することが可能になったからだと考える。

6. おわりに

本稿では時系列方向の学習を行う Time Transformer と時系列データ間関係を学習する Space Transformer を組み合わせることで、他時系列データの情報を予測に生かすことのできるモデルを提案した。現在の Space Transformer は同時刻のみしか時系列間関係を学習できない。今後は、現時点と過去の時点における関係も学習できるようなモデルを作成することで、予測性能を向上させていきたい。

7. 参考文献

[1] Neo Wu, Bradley Green, and Shawn O'Banion. Deep transformer models for time series forecasting: The influenza. CoRR, 2020.

[2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. CoRR, 2017.

^{*1}<https://www.kaggle.com/kannanaikkal/food-demand-forecasting>