

時系列データの将来変動予測における予測根拠の提示

高石 一樹[†] 長尾 智晴[‡]横浜国立大学 大学院環境情報学府[†]横浜国立大学 大学院環境情報研究院[‡]

1. はじめに

近年、人工知能技術の発達により機械学習による時系列予測が盛んになっている[1]。機械学習モデルでは比較的高精度な予測を実現できるものの、なぜその予測を導いたかを説明することは難しく、解釈性は低いといえる。この問題に対し、しばしば線形モデル等の解釈性の高いモデルで元の予測モデルを近似するといった手法が用いられるが、特徴量1つずつの寄与度が得られたとしてもそれらを時系列的なドメイン上で解釈することは難しく、納得感の向上には至らない。本研究では、局所近似手法[2]とトレンドによる予測根拠の提示を組み合わせることで、時系列データの将来変動予測における予測根拠の提示とその解釈を試みた。

2. 提案手法

本研究では、時系列データの将来変動予測における予測根拠の提示手法を提案する。提案手法は任意の時系列予測モデルに適用することができ、出力として「系列Aが上昇したため系列Bが上昇した」といったトレンドベースの説明を提示する。はじめに、学習した任意の予測モデルをSHAP[2]を用いて局所近似する。この時に得られた入力特徴量の寄与度であるshap値の絶対値ベクトル s と入力ベクトル x の要素積を系列 i ごとにとることで(1)、寄与度を考慮した系列毎の入力 x_i' を得る。次に、予め用意したトレンドフィルタ f と x_i' の内積を計算することで(2)、予測に寄与したトレンド度 T_i が系列毎に得られる。また、絶対変動量によって寄与度が左右されるため、(2)で得られたトレンド度 T_i を正規化する(3)。

$$x_i' = s_i \odot x_i \quad (1)$$

$$T_i = x_i' \cdot f \quad (2)$$

$$T_i' = \frac{T_i}{\sum |T_i|} \quad (3)$$

$$f_{up} = (1, 1, 1, 1, 1) \quad (4)$$

$$f_{down} = (-1, -1, -1, -1, -1) \quad (5)$$

表1 データ区間

	Train	Test
海運指数	2010/3/1–2017/5/31	2017/6/1–2018/9/7
原油価格	2012/1/4–2018/8/28	2018/8/29–2019/9/26

3. 実験

3.1 実験設定

本実験は、海運指数と原油価格における1週間後の変動値を予測対象とし、関連系列の過去5日間における変動量を入力、1週間後における自己系列の変動量を出力とする回帰問題である。なお、海運指数の関連系列は自己系列を含む4系列、原油価格の関連系列は自己系列を含む30系列である。また、使用したTrainデータセットとTestデータセットのデータの区間表1に示す。元の予測モデルにはXGBoost[3]を用い、ハイパーパラメータはGrid Searchを用いて探索した。入力は過去2週間分の関連系列の変動値とし、-1.0から1.0に正規化したのちに学習を行う。本手法においてトレンドフィルタには任意のトレンドを抽出するフィルタを設定可能であるが、本実験では5日間の上昇トレンドと下降トレンドを抽出するフィルタ(4)(5)を用いた。

3.2 実験結果

表2に海運指数予測に本手法を適用した際に得られた寄与度を考慮したトレンド度、表3に原油価格予測に本手法を適用した際に得られた寄与度を考慮したトレンド度を示す。また、表4、表5に表2、表3で示した寄与度を考慮したトレンド度を言語で表した結果を示す。

表 2 トレンド度 (海運指数予測)

系列名	上昇	下降
自己系列	0.339	-0.339
FFA1	-0.050	0.050
FFA2	-0.112	0.112
FFA3	-0.498	0.498

表 3 トレンド度(原油価格予測, 上位 3 系列)

系列名	上昇	下降
自己系列	-0.331	0.331
CN Stock Index	-0.133	0.133
No of Transactions	-0.085	0.085

表 4 予測根拠 (海運指数予測)

説明
FFA3 が過去 5 日間で下降したため (0.498)
自己系列が過去 5 日間で上昇したため (0.339)
FFA2 が過去 5 日間で下降したため (0.112)
FFA1 が過去 5 日間で下降したため (0.050)

表 5 予測根拠 (原油価格予測, 上位 3 系列)

説明
自己系列が過去 5 日間で下降したため (0.331)
CN Stock Index が過去 5 日間で下降したため (0.133)
No of Transactions が過去 5 日間で下降したため (0.085)

4. 考察

表 4, 5 より, 予測モデルがどの系列のどのトレンドを予測根拠としているかを出力できていることがわかる. 原油価格予測では入力系列が 30 種あるが, いずれのテストデータに対しても寄与度が高い系列 (0.1 以上) はほぼ 3 系列に留まり, 少数の系列のみで予測を説明できた.

また, 図 1, 2 は表 4 に示した海運指数予測のテストデータに対する実際の自己系列と FFA3 の入力を示す. この図は変動推移を示しており, 背景色の濃さは寄与度の大きさを表している. 図 1 では上昇トレンド, 図 2 では下降トレンドが見られることから, 実際に表 4 で提示されている説明が正しいことがわかる.

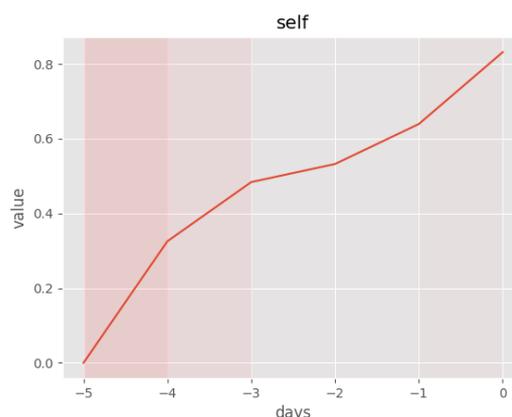


図 1 自己系列の過去 5 日間変動値

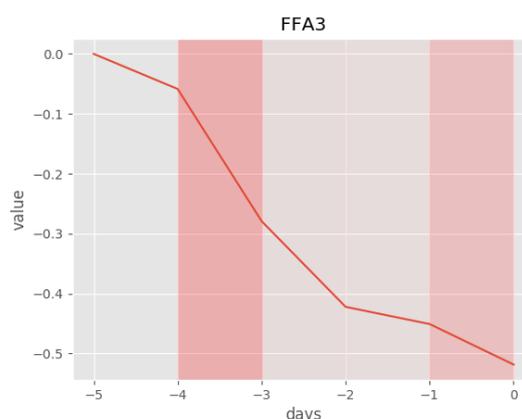


図 2 FFA3 の過去 5 日間変動値

5. おわりに

本研究では局所近似手法とトレンド抽出を組み合わせることで, 時系列データの将来変動予測における予測根拠の提示を試みた. 実験では 1 週間後の海運指数と原油価格の変動量を予測するタスクに本手法を適用し, トレンドをベースとした予測根拠の提示ができることを確認した. 今後は適用対象を増やして検証を進める他, より納得感を得られる説明の提示を目指す.

参考文献

- [1] Abe, Masaya, and Hideki Nakayama. "Deep learning for forecasting stock returns in the cross-section." Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, Cham, 2018.
- [2] Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model predictions." Advances in Neural Information Processing Systems. 2017.
- [3] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. ACM, 2016.