

外生変数を伴う長期予測における主要因子推定法

Evaluation Method of Major Factors in Long-term Prediction with Exogenous Variables

中塚祐喜 *¹

Yuki Nakatsuka

白山晋 *¹

Susumu Shirayama

*¹ 東京大学 工学系研究科

School of Engineering, The University of Tokyo

In many fields, long-term prediction on time series data with exogenous variables has been performed. However, it is difficult to deal with some noise in input data and to qualitatively understand the obtained results. In this paper, first, we extend the DA-RNN proposed by Qin et al so that it can be used in the different time range from the original method. Also, the DA-RNN is extended to be applicable to long-term prediction. Second, we develop a new method of long term prediction based on the extended DA-RNN and the subsequence time-series clustering. An evaluation method of major factors in the exogenous variables is proposed by visualizing the weight of attention mechanism. We tested our method using dataset named SML2010. It is shown that our method has high prediction ability and robustness against noise, accountability of the results.

1. はじめに

1.1 外生変数を伴うデータに対する長期予測

多くの分野・領域において、外生変数を伴う時系列データに対する予測が行われている。その目的は様々であるが、単なる予測でなく、意思決定や政策立案に役立たせるためのものもある。例えば、気温、降水量あるいは降雪量などの予測結果は、商品の仕入れ、電力需要、自然災害対策などに影響する。

外生変数を伴うモデルは、目的変数だけではなく外生変数も参照することで目的変数に対する予測能力を高めている。こうしたモデルの一例として、注意機構を用いた 2 段階注意機構 RNN (Dual-stage Attention-based RNN; DA-RNN) モデル [Qin 17] がある。

このような予測手法は、主に 1 ステップ先の状態を予測する手法として提案されてきた。一方で、近年、時系列予測の分野における新たな課題として、長期予測が注目されている。なお、本稿では長期予測を 2 ステップ以上先までの予測とする。長期予測には、大きく分けて 5 つの予測戦略が提案されている [Taieb 12]。それらの予測戦略は、それぞれ精度の蓄積や確率的従属関係、予測の柔軟性など様々な問題を有している。また、長期予測モデルによって得られる結果は、定量的な結果を与えるものの、人間がすぐに結果を解釈して利用するのは難しく、定性的理解の面で問題がある。

1.2 予測結果の解釈性

上記のように、現状の長期予測における予測戦略は、人間がすぐに結果を解釈して利用するのは難しいという課題がある。解釈性という側面では、近年、機械学習モデルに対する研究でも問題になっている。

Ribeiro らによると、機械学習における解釈性の問題には、「なぜその予測値が得られたのかを説明する」とこと、「そのモデルがどのような性質を持っているかを説明する」とこという 2 つの観点がある [Ribeiro 16]。

本稿で扱う注意機構を利用したモデルは後者に分類することができる。注意機構は、深層学習の入力層や中間層のベクトルから予測に有用な部分を自動的に学習する方法であり、機械翻

訳の分野を中心に既存のモデルの予測能力を高めるものとして注目されている。さらに、注意機構では、注意機構の重みを可視化することで与えられた入力の中の部分が有用であったかを知ることができる。Riemer らは目的変数と外生変数からなる入力に対して、注意機構を利用した解釈性の高い時系列モデルを提案している [Riemer 16]。彼らの研究では、予測における主要因子の重要度を決定することに注意機構を利用している。

1.3 本研究の目的

以上に述べた時系列データに対する解釈性の高い注意機構モデルには、長期予測に対して提案されたものはない。一方で、応用上、長期予測モデルが必要となるケースは多く存在する。また、外生変数の時系列が部分的に目的変数と相関があるように振る舞うことがあり、既存のモデルでは目的変数と本来関係のない誤った入力変数を選択する可能性がある。

そこで、本研究の目的を外生変数を伴う時系列データに対する長期予測タスクにおける解釈性の高い注意機構モデルを提案することとし、同時に注意機構の重みをより解釈しやすい方法を提案する。加えて、部分的なノイズにも頑健なモデルを提案することを目標とする。

2. 外生変数を伴う予測モデル

本節では外生変数を伴う時系列モデルとして Dual-stage Attention-based RNN (DA-RNN) モデルを説明する。

本稿では以下の 2 通りの定式化を扱う。

$$\hat{y}_{T+1} = F(y_1, y_2, \dots, y_T, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T) \quad (1)$$

$$\hat{y}_{T+1} = F(y_1, y_2, \dots, y_T, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T, \mathbf{x}_{T+1}) \quad (2)$$

ここで、 y_t ($t = 1, 2, \dots, T$) は目的変数、 \mathbf{x}_t ($t = 1, 2, \dots, T$) は外生変数を表す。また、以下では、外生変数の数を n とし、外生変数を $\mathbf{x}^k = (x_1^k, \dots, x_{T+1}^k)$ ($k = 1, 2, \dots, n$) と表す。

2.1 DA-RNN モデル

DA-RNN モデルは入力情報・時系列情報の両方に注意機構を組み込んだモデルである。入力情報における注意機構を第 1 段階の注意機構、時系列情報における注意機構を第 2 段階の注意機構と呼ぶ。DA-RNN における Encoder と Decoder は、LSTM のユニットで構成されている。

連絡先: 中塚 祐喜, 東京大学大学院工学系研究科システム創成学専攻 白山研究室, 文京区本郷 7-3-1, nk.tsssa@gmail.com

2.1.1 第1段階の注意機構

第1段階の注意機構は、Encoderにおける注意機構である。時点 $t-1$ での Encoder の m 次元の隠れ状態ベクトル \mathbf{h}_{t-1} 、内部セルの状態ベクトル \mathbf{s}_{t-1} に対し、次時点 t 、特徴 k における関連度 e_t^k を

$$e_t^k = \mathbf{v}_e^T \tanh(\mathbf{W}_e[\mathbf{h}_{t-1}; \mathbf{s}_{t-1}] + \mathbf{U}_e \mathbf{x}^k) \quad (3)$$

と計算する。なお、 $\mathbf{v}_e \in \mathbb{R}^T$ 、 $\mathbf{W}_e \in \mathbb{R}^{T \times 2m}$ 、 $\mathbf{U}_e \in \mathbb{R}^{T \times T}$ であり、これらは学習するパラメータである。

そして、この関連度 e_t^k をソフトマックス層に通すことで、注意機構の重み α_t^k を計算する。この α_t^k を用いて $\tilde{\mathbf{x}}_t = (\alpha_t^1 x_t^1, \dots, \alpha_t^n x_t^n)$ とし、Encoderに入力することで、最終的な隠れ状態ベクトル \mathbf{h}_t を得る。

2.1.2 第2段階の注意機構

第2段階の注意機構は、Decoderにおける注意機構である。時点 $t-1$ での Decoder の m 次元の隠れ状態ベクトル \mathbf{d}_{t-1} 、内部セルの状態ベクトル \mathbf{s}'_{t-1} に対し、時点 t の第2段階の注意機構における関連度 l_t^i を

$$l_t^i = \mathbf{v}_d^T \tanh(\mathbf{W}_d[\mathbf{d}_{t-1}; \mathbf{s}'_{t-1}] + \mathbf{U}_d \mathbf{h}_i) \quad (4)$$

と計算する。ここで、 $\mathbf{v}_d \in \mathbb{R}^m$ 、 $\mathbf{W}_d \in \mathbb{R}^{m \times 2m}$ 、 $\mathbf{U}_d \in \mathbb{R}^{m \times m}$ であり、これらは学習するパラメータである。

そして、この関連度 l_t^i をソフトマックス層に通し、注意機構の重み β_t^i を得る。この β_t^i を用いて、Encoderの重み付き隠れ状態ベクトル \mathbf{c}_t を $\mathbf{c}_t = \sum_{i=1}^T \beta_t^i \mathbf{h}_i$ のように得る。こうして得た隠れ状態ベクトルを Decoderに入力する。

Decoderでは、隠れ状態ベクトル \mathbf{c}_t と過去の目的変数 y_t を入力として受け取り、隠れ状態ベクトル \mathbf{d}_t を計算する。この \mathbf{d}_t を出力層に通し、次のように予測値 \hat{y}_{T+1} を得る。

$$\hat{y}_{T+1} = \mathbf{v}_y^T (\mathbf{W}_y[\mathbf{d}_{T+1}; \mathbf{c}_{T+1}] + \mathbf{b}_w) + b_v \quad (5)$$

ただし、 $[\mathbf{d}_{T+1}; \mathbf{c}_{T+1}] \in \mathbb{R}^{2m}$ である。また、 $\mathbf{W}_y \in \mathbb{R}^m \times 2m$ 、 $\tilde{\mathbf{b}}_w \in \mathbb{R}^m$ 、 $\mathbf{v}_y \in \mathbb{R}^m$ 、 $b_v \in \mathbb{R}$ であり、これらは学習するパラメータである。

2.1.3 DA-RNN モデルの限界

DA-RNN は式 (2) の形で定式化されており、将来の時間 $t+1$ 時点での外生変数の値を一般には利用できない点が問題であると指摘されている [Tao 18]。また、この手法では、時系列が一部の区間でのみ相関があるように振る舞った場合、誤った入力変数を選択する可能性があると考えられる。

3. 提案手法

3.1 DA-RNN モデルの拡張

DA-RNN の元の定式化は、式 (2) の形式であった。そこで本稿では、Tao の指摘を踏まえて、DA-RNN を式 (1) に合わせた定式化を提案する。具体的には、式 (5) の計算を次のように変更する。

$$\hat{y}_{T+1} = \mathbf{v}_y^T (\mathbf{W}_y[\mathbf{d}_T; \mathbf{c}_T] + \mathbf{b}_w) + b_v \quad (6)$$

また、Encoder の隠れ状態ベクトルは次のように更新する。

$$\mathbf{h}_t = f_1(\mathbf{h}_{t-1}, \tilde{\mathbf{x}}_{t-1}) \quad (7)$$

この変更に合わせて、Decoder の隠れ状態ベクトルは時点 T まで計算する。本研究では、この方法を DA-RNN-Tp モデル (DA-RNN-Tp と表記) と呼ぶことにする。

3.2 長期予測モデルへの拡張

次に、長期予測への拡張法を提案する。ここで、長期予測とは τ ステップ先までの予測を指す。本稿では、予測値 \hat{y}_{T+1} を得た後、その予測値を再び Decoder に入力することで、次のステップの予測値 \hat{y}_{T+2} を計算する。この操作を τ 回繰り返すことで長期予測を行う。

この方法によって、DA-RNN-Tp を長期予測に適用できるようにする。本研究では、このモデルを DA-RNN-TpLP と記す。

3.3 部分時系列クラスタリング付きの DA-RNN-TpLP

前節までに示した DA-RNN モデルの拡張版に対し、部分時系列クラスタリングによるクラスタの情報をモデルに加えることによって、よりノイズに頑健なモデルを構築する。具体的には、前処理として訓練データにおける外生変数と目的変数の過去の値を全てを使って部分時系列クラスタリングを行う。スライディングウィンドウによって D 個の訓練データが得られた場合、つまり、 $((\mathbf{x}^1, \dots, \mathbf{x}^n, \mathbf{y}), y_{T+1})$ の組が D 個ある場合、 $(n+1) \times D$ 個の部分時系列を用いて部分時系列クラスタリングを行う。

ここでは、部分時系列クラスタリング手法として k-shape 法 [Paparrizos 15] と呼ばれる手法を利用する。

以下では、k-shape 法によって得られた $\mathbf{x}^1, \dots, \mathbf{x}^n, \mathbf{y}$ に対応するクラスタを $\mathbf{c}_1, \dots, \mathbf{c}_n, \mathbf{c}_{n+1}$ と表す。また、これらを one-hot ベクトルによって表したものを $\mathbf{c}_1, \dots, \mathbf{c}_n, \mathbf{c}_{n+1}$ とする。

このように得られたクラスタ情報を用いて外生変数のクラスタ $\mathbf{c}_1, \dots, \mathbf{c}_n$ と、目的変数のクラスタ \mathbf{c}_{n+1} を比較し、目的変数と類似したクラスタ同士の重みが大きくなるように、式 (3) の e_t^k を次のように補正する。

$$e_t'^k = e_t^k (U^k (\mathbf{c}_k - \mathbf{c}_{n+1})) \quad (8)$$

ただし、 $U^k \in \mathbb{R}^C$ であり、学習パラメータである。ここで C はクラスタの総数である。以降の計算は DA-RNN-TpLP と同様である。本稿では、この方法を提案手法と称する。

3.4 注意機構の重みによる主要因子推定法と情報支援

最後に、注意機構の重みを利用した新たな主要因子推定法を提案する。提案手法や DA-RNN では、第1段階の注意機構によって、各データに対し各特徴における注意機構の重みを計算することができる。この注意機構の重みの平均や分散によってその特徴を記述し、他の特徴と比較することも可能になる。しかしながら、平均や分散が近い値をとる場合でも、各特徴によって注意機構の重みの分布の形が異なる可能性がある。このため、注意機構の重みの分布を比較する方法が必要になる。

本稿では Bhattacharyya 距離 [Bhattacharyya 43] を利用して、特徴をクラスタリングすることを提案する。クラスタリング方法として、k-means++法を利用する。クラスタリングによって得られたあるグループに対し、得られた特徴と目的変数を同時にプロットすることで、それをユーザへの支援情報として活用することができる。その結果、そのような支援情報に基づき必要な特徴の取捨選択や特徴の追加などを考察することが可能となる。

4. 数値実験

本稿では、数値実験に SML2010 データセット [Zamora-Martínez 14] を用いる。SML2010 は室温

表 1: 短期予測における評価指標

	RMSE	MAE	MAPE
Encoder-Decoder	2.800	2.096	0.100
Attention	2.608	1.971	0.095
DA-RNN-Tp	2.572	1.886	0.091
Proposed	2.508	1.812	0.088

表 2: 長期予測における評価指標

	RMSE	MAE	MAPE
DirRec	0.252	0.186	0.009
Encoder-Decoder	0.249	0.183	0.009
Attention	0.281	0.216	0.010
DA-RNN-TpLP	0.233	0.157	0.007
Proposed	0.239	0.164	0.008

表 3: ノイズを追加した場合の長期予測における評価指標

	RMSE	MAE	MAPE
DA-RNN	0.556	0.361	0.017
Proposed	0.555	0.352	0.016

表 4: ノイズを追加した場合の短期予測における評価指標

	RMSE	MAE	MAPE
DA-RNN	0.051	0.039	0.002
Proposed	0.048	0.039	0.002

の予測に利用されるデータセットであり、屋内外に設置された温度計・湿度計などのセンサの値を収集したものである。目的変数として室温のデータを、外生変数として屋内外の湿度・風速などの $n = 20$ 個の外生変数を用いる。

これらのデータは合計約 40 日間、1 分ごとに計測されているが、15 分の区間幅で平滑化されている。本実験では、訓練データとして最初の 3,200 点、検証データとして続く 400 点、テストデータとして最後の 537 点を利用する。検証データは学習の終了エポック数の決定に利用した。入力データのウィンドウ幅 $T = 10$ 、隠れ状態ベクトルの次元数 $m = p = 128$ とし、クラスタ数 $C = 8$ とし、損失関数には平均二乗誤差 (MSE) を利用した。そして、短期予測・長期予測 2 種類のタスクにおいて手法の比較を行った。長期予測では、 $\tau = 10$ ステップ先までの予測を行った。

上記の設定の下、手法の性能比較を行うために RMSE, MAE, MAPE という 3 つの評価指標を採用した。

4.1 実験結果

短期予測においては 3 種類の従来手法とその性能を比較した。使用した従来手法は、Encoder-Decoder モデル、注意機構モデル、DA-RNN-Tp モデルである。長期予測においては上記 3 種類のモデルに加え、DirRec Strategy [Sorjamaa 06] と LSTM を組み合わせたモデルとも比較を行った。

表 1, 表 2 にテストデータに対する評価指標の値を示す。

短期予測の場合、表 1 に示すように、DA-RNN-Tp と提案手法は他の比較手法よりも優れた結果を示した。さらに、提案手法は全ての指標で DA-RNN-Tp よりも優れた予測能力を示した。これは部分時系列クラスタリングによって、効率的に注意機構の重みを計算することが可能であることを示唆している。

長期予測の場合、短期予測よりも予測区間が長くなるため、全体として予測能力が落ちているものの、2 段階の注意

機構のモデルの予測精度の高さが確認された。中でも、DA-RNN-TpLP が全ての評価指標において他の全ての手法よりも優れた予測能力を示した。

4.2 注意機構の重みの比較と特徴の分析

さらに、DA-RNN-Tp と提案手法における外生変数の注意機構の重みの平均値を計算した。SML2010 における結果を図 1 に示す。ただし、Bhattacharyya 距離 [Bhattacharyya 43] を距離関数とした際のクラスタリングによって同じクラスタに割り当てられた特徴を同じ色で示している。クラスタ数は 5 であるので、全部で 5 色に各特徴が分けられている。

図から、両者の注意機構の重みの傾向は類似しているものの、比較手法よりも提案手法の方が外生変数における注意機構の重みの大小が明確になっていることがわかる。したがって、提案手法は従来手法よりも不要な外生変数を自動的に選別して予測を行っていると考えられる。

また、各特徴に割り当てられた特徴に関して、注意機構の重みの平均が近いもの同士が同じクラスタに割り当てられている傾向にある。しかしながら、例えば「Precipitacion」という特徴と「Meteo.Exterior.Sol.Oest」という特徴は重みの平均値差は小さいにも関わらず、重みの分布が大きく異なっているため、異なったクラスタが割り当てられている。図 2 にこれらの特徴の注意機構の重みの分布を示す。図のように重みの分布に大きな差異が見られており、モデルにその特徴が大きな重要性を持つと判断される機会に違いがあることがわかる。

4.3 ノイズに対する頑健性の比較

提案手法のノイズに対する頑健性を調べるために、既存の外生変数に対し一定のノイズを加える実験と全く予測に寄与しないノイズに基づく特徴を加える実験を行った。

4.3.1 一部区間におけるノイズ

この実験では、標準化したテストデータに対し、正規分布に従うノイズを加えてその予測能力の変化を調べた。表 3 に長期予測における結果を示す。表 2 と表 3 を比較すると、ノイズがない場合には DA-RNN-TpLP が優れた結果を示していたが、ノイズを加えた場合にはその予測能力が逆転した。この実験結果は、提案手法は部分的なノイズに対してより頑健であることを示唆している。

4.3.2 入力変数におけるノイズ

この実験では、全く予測に寄与しないノイズに基づく特徴を外生変数として追加して、提案手法における予測能力や注意機構の重みの変化を調べた。

正規分布からサンプリングしたノイズに基づく特徴を 5 種類追加し、合計 25 個の外生変数で短期予測モデルを訓練し、その予測性能をテストデータで評価を行ったところ、RMSE: 0.026, MAE: 0.019, MAPE: 0.001 という結果を得た。表 1 と結果を比較すると、わずかな予測能力の劣化は見られるが、ほとんどノイズの影響を受けていないと考えられる。また、図 3 はこの時の注意機構の重みを示したものである。図より、追加したノイズに対しては一貫して低い注意機構の重みを与えていることが確認できる。

5. おわりに

本稿では、長期予測モデルに対する解釈性の高いモデルが存在しないことと既存の予測モデルは部分的なノイズに対して頑健ではないことを指摘した上で、Qin らの DA-RNN モデルを長期予測が可能になるように拡張し、部分時系列クラスタリングを加えることで、ノイズに対して頑健性を有する長期予測

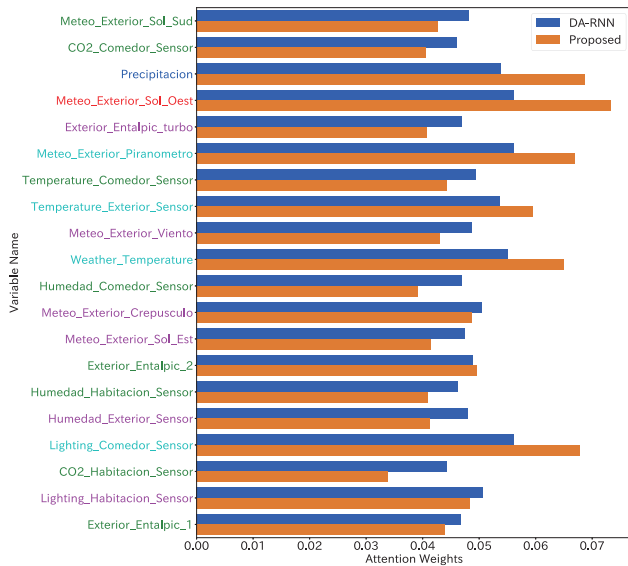


図 1: SML2010 における短期予測の場合の訓練データの注意機構の重み

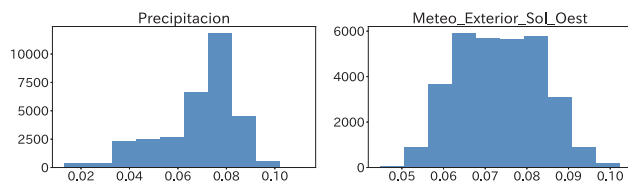


図 2: 注意機構の重みの分布の比較

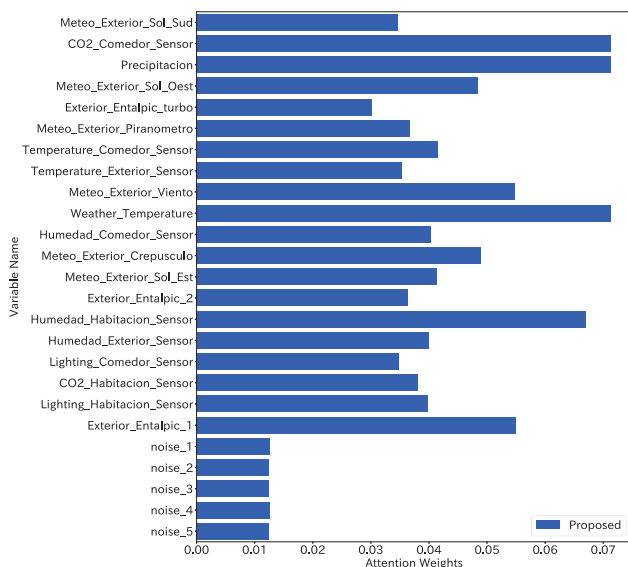


図 3: ノイズ特徴を加えた場合の注意機構の重み

手法を構築した。また、その注意機構の重みから主要因子を推定する方法を提案した。

提案手法を複数のデータセットに適用し、短期・長期予測の両方において従来手法と同等以上の予測能力があることを示した。また、ノイズを加えた場合でも高い予測能力を維持できることから、ノイズに対して頑健であることを示した。

今後の課題としては、注意機構の重みの解釈性を向上させる可視化方法の開発、部分時系列クラスタリングのクラスタ数の適切な決定方法の考案、可視化における解釈性の高いクラスタ数の決定方法の開発などが挙げられる。

参考文献

- [Bhattacharyya 43] Bhattacharyya, A.: On a measure of divergence between two statistical populations defined by their probability distributions, *Bulletin of the Calcutta Mathematical Society*, Vol. 35, pp. 99–109 (1943)
- [Paparrizos 15] Paparrizos, J. and Gravano, L.: k-Shape: Efficient and accurate clustering of time series, in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pp. 1855–1870 (2015)
- [Qin 17] Qin, Y., Song, D., Cheng, H., Cheng, W., Jiang, G., and Cottrell, G. W.: A dual-stage attention-based recurrent neural network for time series prediction, in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pp. 2627–2633 (2017)
- [Ribeiro 16] Ribeiro, M. T., Singh, S., and Guestrin, C.: Why should I trust you?: Explaining the predictions of any classifier, in *Proceedings of the 22nd International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144 (2016)
- [Riemer 16] Riemer, M., Vempaty, A., Calmon, F., Heath, F., Hull, R., and Khabiri, E.: Correcting forecasts with multifactor neural attention, in *Proceedings of International Conference on Machine Learning*, pp. 3010–3019 (2016)
- [Sorjamaa 06] Sorjamaa, A. and Lendasse, A.: Time series prediction using DirRec strategy., in *Proceedings of the European Symposium on Artificial Neural Networks*, pp. 143–148 (2006)
- [Taieb 12] Taieb, S. B., Bontempi, G., Atiya, A. F., and Sorjamaa, A.: A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition, *Expert Systems with Applications*, Vol. 39, No. 8, pp. 7067–7083 (2012)
- [Tao 18] Tao, Y., Ma, L., Zhang, W., Liu, J., Liu, W., and Du, Q.: Hierarchical Attention-Based Recurrent Highway Networks for Time Series Prediction, *arXiv preprint arXiv:1806.00685* (2018)
- [Zamora-Martínez 14] Zamora-Martínez, F., Romeu, P., Botella-Rocamora, P., and Pardo, J.: On-line learning of indoor temperature forecasting models towards energy efficiency, *Energy and Buildings*, Vol. 83, pp. 162–172 (2014)