

深層学習による各家庭の給湯消費電力量の予測:
 Uncertainty Weighting による分類・回帰のマルチタスク学習
 Prediction of Domestic Electric Power Consumption in Water Heater:
 Multi-Task Learning with Uncertainty Weighting

新美 潤一郎[†] 塚本 高浩[‡] 竹内 真登[§] 柴田 与夢[¶]
 Junichiro Niimi Takahiro Tsukamoto Makito Takeuchi Atomu Shibata

1. はじめに

近年、生活様式の多様化や、オール電化等の導入、電気自動車の普及、さらにはガスに電力を組み合わせたハイブリッド給湯器の導入などの要因により、各家庭における電力消費パターンの多様化が進んでいる。必ずしも従来の手法のみでは十分に予測ができない状況となっていることから、高精度の予測モデルの開発は急務となっている。これは需要予測のみならず、ピークシフトや平準化などのエネルギーマネジメントにも直結する重要な課題である。一方で、近年の機械学習分野、特に深層学習手法の飛躍的な発展により、エネルギー分野においてもさまざまな手法を用いた予測モデルの提案が続いている [1, 2]。それらの手法の多くは非線形活性化と階層性により、従来の統計手法と比較しても高い予測性能を有しているものの、高頻度で取得される IoT データへの応用研究は未だ限定的であり、実務応用のための高精度な手法の導入にはさらなるモデルの改良が必要である。そこで本研究では、ガス/電力のハイブリッド給湯器が収集する IoT ログから 30 分単位の給湯電力の消費量を予測する深層学習モデルを提案する。

2. 関連研究

まず、本研究の位置付けを確認するために関連研究の整理を行う。

2.1 電力消費量の予測

家庭における電力消費量の予測はさまざま行われており、Multi-Layer Perceptron (MLP) [3] をはじめとした深層学習を用いることの有用性 [1, 2] が示されてきた。また、時系列解析一般においては ARIMA 等の統計手法と比較して LSTM が高い精度を発揮することも指摘されている [4]。近年では複数のタスクを同時並行的に最適化するマルチタスク学習 (MTL) [5] が進んでおり、特に各タスクの損失関数の分散を不確実性パラメータとして最適化する uncertainty weighting (UW) [6] が提案されている。UW では画像認識モデルを semantic タスク、instance タスク、depth タスクに分岐させ、各損失関数を UW により最適化した上で、全出力を統合して最終的な予測を行う構造をとっている。

本研究では予備的な解析により、電力が消費されていない間でもモデルが正確に消費量 0W を予測することが困難という課題が見出されている。そこで、UW により一般的な消費量の予測と同時に、消費の有無に関する離散的な予

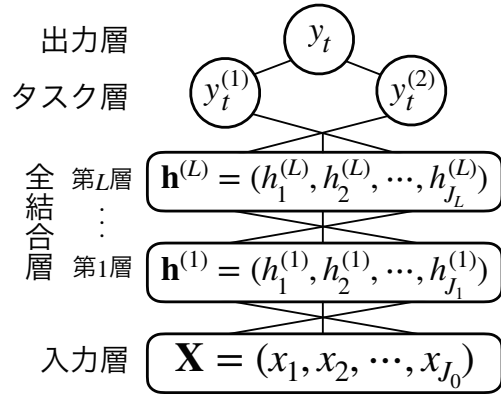


図 1 Model architecture (MTL-MLP)

測を行い予測値を補正することにより、従来手法より高い精度での予測を実現できる可能性がある。

3. 提案モデル

次に本研究の提案モデルの構造と最適化の方法、そしてモデルの評価方法を示す。

3.1 モデル構造

本研究では 1 日を 48 枠に分割し、各単位時間における給湯電力の消費量の予測を世帯レベルで行う。サンプルサイズ n のデータセットにおいて、あるサンプル $i \in \{1, 2, \dots, n\}$ において、ミニバッチ学習のバッチサイズ $bs \in \{1, 2, \dots, n\}$ とすれば、あるミニバッチ $k \in \{1, 2, \dots, \lceil n/bs \rceil\}$ における標本空間 $\mathcal{B}_k = \{i \mid (k-1)bs + 1 \leq i \leq \min(k \cdot bs, n)\}$ となる。

図 1 に示す MLP の隠れ層の数を $L > 1$ 、ある隠れ層 $l \in \{1, 2, \dots, L\}$ における次元数を dim_l とすれば、MLP の最後の隠れ層の状態は $\mathbf{h}^{(L)} \in \mathbb{R}^{bs \times dim_L}$ である。ここで分散タスクと回帰タスクそれぞれを示すインディケータを $m = \{1, 2\}$ 、タスク m への出力用パラメータとして、ウェイト $\mathbf{W}^{(m)} \in \mathbb{R}^{dim_L \times 1}$ 、バイアス $\mathbf{b}^{(m)} \in \mathbb{R}^{dim_L}$ 、活性化関数 $\phi^{(m)}$ を用いて、

$$\hat{y}_k^{(m)} = \phi^{(m)}(\mathbf{h}_k^{(L)} \mathbf{W}^{(m)} + \mathbf{b}^{(m)}) \in \mathbb{R}^{bs \times 1} \quad (1)$$

である。ここで活性化関数として用いるのはそれぞれシグモイド関数 [3]: $\phi^{(1)}(u) = 1/(1 + e^{-u})$ および正規化線形ユニット (ReLU) [7]: $\phi^{(2)}(u) = \max(0, u)$ である。

最終的な出力の算出にあたり、分類タスクの予測値を離散化する必要がある。ある Epoch $t \in \{1, 2, \dots, T\}$ における離散化の閾値を $thr_t \in [0, 1]$ とすれば、

$$\hat{y}_k^{(1)} = 1[\hat{y}_k^{(1)} > thr_t] \in \{0, 1\}^{bs \times 1} \quad (2)$$

[†] 名城大学, Meijo University; 理化学研究所, RIKEN AIP

[‡] 中京大学, Chukyo University

[§] 名古屋市立大学, Nagoya City University

[¶] リンナイ, Rinnai Corporation

で分類タスクの予測値を離散化できる。よって最終的な予測値は、タスク固有の予測値のアダマール積 \odot を用いて、

$$\hat{y}_k = \tilde{y}_k^{(1)} \odot \hat{y}_k^{(2)} \quad (2)$$

と表すことができる。

3.2 最適化

次に、モデルの最適化として、本研究では離散タスクと回帰タスクそれぞれについて、binary cross-entropy (BCE) [8] と log mean square error (Log-MSE) [8] の最小化を行う。ここで $\mathbf{1}_{bs}$ を全要素が1である長さ bs の列ベクトルとして $\epsilon = 10^{-8} \mathbf{1}_{bs}$ を導入すると、あるミニバッチ k におけるタスク m の損失関数 $\mathcal{L}_k^{(m)}$ はそれぞれ、

$$\mathcal{L}_k^{(1)} = -\frac{1}{|\mathcal{B}_k|} \left[\mathbf{y}_k^{(1)T} \log(\hat{\mathbf{y}}_k^{(1)} + \epsilon) + (1 - \mathbf{y}_k^{(1)})^T \log(1 - \hat{\mathbf{y}}_k^{(1)} + \epsilon) \right] \quad (3)$$

$$\mathcal{L}_k^{(2)} = \frac{1}{|\mathcal{B}_k|} \left\| \log(\hat{\mathbf{y}}_k^{(2)} + \epsilon) - \log(\mathbf{y}_k^{(2)} + \epsilon) \right\|_2^2 \quad (4)$$

である¹。 ϵ はlogの適用にあたって数値の発散を防ぐことを目的に加算している。最後に、タスク固有の損失関数の分散に基づく不確実性パラメータ $\nu^{(m)}$ を導入したUWにより、モデル全体を包括する損失関数 \mathcal{L}_{global} を、

$$\mathcal{L}_{global} = \sum_{m=1}^2 \left[\frac{1}{2\nu^{(m)}} \mathcal{L}_k^{(m)} + \log(\nu^{(m)}) \right] \quad (5)$$

とする。これら不確実性パラメータ自体も勾配降下法により更新されることにより、規模の異なる複数タスクの損失関数をスケールした単一の損失の最適化が可能である。

また、離散化の閾値 thr の最適化は、黄金比探索 [9] により学習セット内で各エポックにF1値が最も向上した値を採用する。

3.3 評価方法

最後に、提案モデルの有効性を示すにあたっては評価指標としてroot mean square error (RMSE)を用いる。また、主要なベースラインとして深層学習としてTransformer [10]、LSTM [11]、MLPを、機械学習としてXGBoost [12]とLinear SVM [13]を、最後に時系列的な統計手法としてARIMAX [14]で予測モデルを作成し精度を比較する。

4. 解析

解析にあたっては、各世帯に設置された給湯器から収集されたIoTログを用いる。過去30日間の消費状況から翌1日の48枠それぞれの消費電力量を予測する。得られたサンプルサイズ $n = 180,805$ から、テストセットへのdata leakageを防ぐため世帯レベルで学習・検証・テストにそれぞれ70%, 15%, 15%を割り当てた上で、世帯によってデータ取得期間が異なることへの対応として、各世帯から単位時間レベルでサンプルを抽出する。

¹ 式(4)は数式の簡易化のためl2 normの2乗による表記としているが、実際のモデル構築では浮動小数の計算上の限界から平方根と累乗を繰り返すことで結果が変化するため、MSEを直接的に計算している。

表1 Results (Lift, validation-best)¹

Model Name	Task	Train	Val	Test
MLP	MTL	0.606	0.658	0.703
Transformer	single	0.604	0.732	0.732
LSTM	single	0.797	0.699	0.734
MLP	single	1.000	0.907	0.907
XGBoost	single	0.072	0.967	0.960
ARIMAX	single	0.898	1.204	1.209
Linear SVM	single	1.604	1.915	1.682

表1に単一タスクのMLPを基準とした精度の相対値を示すと、学習時にはXGBoostが、検証/テスト時には提案モデルが最も高い精度を示した。主な傾向として、TransformerやLSTM等、深層学習の発展的な手法を中心に予測性能が改善している一方で、単一タスクのMLPでは、伝統的な機械学習や時系列手法よりは精度が高いものの、相対的には性能が必ずしも向上しない結果となっている。提案モデルでは学習時の精度はTransformerとほとんど変わらないものの、検証・テストにおける汎化性は大きく上回っている。よって、今後はTransformer等のMTL拡張により更なるモデル改善の可能性がある。

謝辞

本研究は科研費(24K16472, 25K00678)の支援を受けています。

参考文献

- [1] Frikha Majdi, Taouil Faouzi, Fakhfakh Ahmed, Derbel Faouzi, "Predicting power consumption using deep learning with stationary wavelet", *Forecasting*, Vol.6, No.3 (2024).
- [2] Kong Weicong, Dong Yang Zhao, Jia Youwei, Hill J. David, Xu Yan, Zhang Yuan, "Short-term residential load forecasting based on LSTM recurrent neural network", *IEEE Transactions of Smart Grid*, Vol.10, No.1 (2017).
- [3] Rumelhart E. David, Hinton E. Geoffrey, Williams J. Ronald, "Learning representations by back-propagating errors", *Nature*, Vol.323, No.6088 (1986).
- [4] Siami-Namini Sima, Tavakoli Neda, and Namin Siami Akbar "A comparison of ARIMA and LSTM in forecasting time series", *2018 17th IEEE international conference on machine learning and applications* (2018).
- [5] Caruana Rich, "Multitask learning", *Machine Learning*, Vol.28 (1997).
- [6] Kendall Alex, Gal Yarin, Cipolla Roberto, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics", *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018).
- [7] Glorot Xavier, Bordes Antoine, Bengio Yoshua, "Deep Sparse Rectifier Neural Networks", *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, Vol.323, No.6088 (1986).
- [8] Murphy P. Kevin, *Machine Learning: A Probabilistic Perspective* (2012).
- [9] Kiefer Jack, "Sequential minimax search for a maximum", *Proceedings of the American mathematical society*, Vol.4, No.3 (1953).
- [10] Vaswani Shazeer Noam, ..., Polosukhin Illia "Attention is all you need", *Advances in neural information processing systems* Vol.30 (2017).
- [11] Hochreiter Sepp, Schmidhuber, Jürgen "Long short-term memory", *Neural computation* Vol.9, No.8 (1997).
- [12] Chen Tianqi, Guestrin Carlos "XGBoost: A Scalable Tree Boosting System", *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016)
- [13] Cortes Corinna, Vapnik Vladimir "Support-Vector Networks", *Machine learning*, Vol.20 (1995).
- [14] Box E. P. George, Jenkins M. Gwilym, *Time Series Analysis: Forecasting and Control* (1970).