

Flare Transformer Regressor:
Masked Autoencoder と Informer Decoder に基づく太陽フレア予測
Flare Transformer Regressor:
Solar flare prediction based on Masked Autoencoder and Informer Decoder

九曜 克之¹⁾ 和田 唯我¹⁾ 兼田 寛大¹⁾ 飯田 紡¹⁾
Katsuyuki Kuyo Yuiga Wada Kanta Kaneda Tsumugi lida
西塚 直人²⁾ 久保 勇樹²⁾ 杉浦 孔明¹⁾
Naoto Nishizuka Yûki Kubo Komei Sugiura

1 はじめに

太陽フレアによって放出される X 線や高エネルギー粒子などは、電波障害や停電、宇宙飛行士への健康的被害などをもたらす [1]. また、Carrington 級フレアによる経済的損失は北米で約 1,630 億ドルと推定されている [20]. そのため、宇宙天気予報による太陽フレアの予測技術が重要である。しかし、太陽フレアの発生を正確に予測することは非常に難しい。

専門家でも太陽フレアを予測するのは非常に困難である。例えば、2000–2015 年の期間における専門家予測は $GMGS = 0.48$ (Gandin–Murphy–Gerrity score) および $TSS_{\geq M} = 0.50$ (true skill statistics) であった [8]. ただし、これらの指標は完璧な予測に対してスコア 1 を返す。さらに、専門家による太陽フレアの回帰予測は行われていない。

太陽フレア予測の分野において、専門家による予測は時刻 t から 24 時間以内の最大の太陽フレアクラスを求める分類問題として取り組まれている。既存手法においても、専門家による予測と比較するために分類問題として扱うことが標準的な設定であった (例: [12] [15]). Flare Transformer [25] は、太陽フレア予測を分類問題として取り組み、専門家予測を超える性能を達成した。そのため、今後は専門家にとって難しい、太陽フレアの X 線強度に対する回帰予測が重要になるとと思われる。そこで、本研究では太陽フレア予測を回帰問題として扱う。

このような背景から、本論文では、時刻 t から 24 時間後までの太陽フレアの X 線強度を予測するタスクを扱う。本タスクでは、時刻 $t-k+1$ から時刻 t までの 1 時間おきの物理特徴量と磁場画像を入力とし、時刻 $t+1$ から時刻 $t+24$ までの太陽フレアの X 線強度を予測する。

物理特徴量と磁場画像を用いた Flare Transformer [25] は太陽フレア予測タスクにおいて専門家による予測を超える性能を達成した。しかし、Flare Transformer は、24 時間以内において最大の太陽フレアクラスの発生確率を予測する分類問題を扱っており、太陽フレアの X 線強度を直接予測する回帰問題は扱っていない。また、Flare Transformer における Image Feature Extractor は表現力が低く、磁場画像から有用な特徴量を十分に抽出できていない可能性がある。

以上のことから、本論文では、磁場画像と物理特徴量を入力とし、時刻 t から 24 時間後までの太陽フレアの

X 線強度の予測を行う Flare Transformer Regressor を提案する。既存手法との違いは、太陽フレア予測を回帰問題として扱うために、Informer [24] を参考にした回帰用デコーダを構築した点である。

提案手法では、太陽フレア予測に寄与する特徴量を抽出するため、Masked Autoencoder [6] を用いて事前学習を行った Vision Transformer (ViT) [3] を導入する。また、Informer [24] を参考にした回帰用デコーダを構築することで、太陽フレア予測に重要な特徴量に注目し、長時間の時系列予測を効率的に行うことができると考えられる。

本研究の独自性は以下である。

- 太陽フレア予測に寄与する特徴量を抽出するため、Masked Autoencoder で事前学習した ViT を導入する。
- 太陽フレア予測を回帰問題として扱うために、Informer [24] を参考にした回帰用デコーダを構築する。

2 関連研究

時系列予測の研究は盛んに行われている [23] [24]. サーベイ論文である [4] では、時系列予測における各手法と既存データセット、またサブタスクに関する総括がなされている。

時系列予測分野においては、様々な手法が提案されている。古典的な手法として、統計モデルに基づく手法 [2] [7] がある一方、近年では DNN (Deep Neural Networks) を用いた様々な手法が提案されている。DNN による既存手法は二つに大別される。すなわち、RNN (Recurrent Neural Network) ベースの手法 [14, 17, 18, 22], および Attention ベースの手法 [23] [24] [10] である。例えば RNN ベースである DeepAR [18] では、将来における事象の確率分布を自己回帰型 RNN で予測する一方、Attention ベースである Informer [24] では、ProbSparse self-attention 機構を用いた注意機構により、生成的デコーダを通して将来の時系列を予測する。

太陽フレアの予測についても、これまでに多くの研究がなされている [12] [15] [9]. [5] では、太陽フレア予測における 14 種類の機械学習手法を含む手法や評価指標についての包括的な総括がなされている。

DeFN [12] は 24 時間以内に太陽フレアが発生するかどうかを黒点領域ごとに予測するモデルである。この手法では物理特徴量の各要素について出力への寄与度を分析することができるという利点がある。また DeFN は決定論的な予測を行う一方、DeFN を拡張した DeFN-R [13] で

1) 慶應義塾大学 Keio University

2) 国立研究開発法人情報通信研究機構 National Institute of Information and Communications Technology

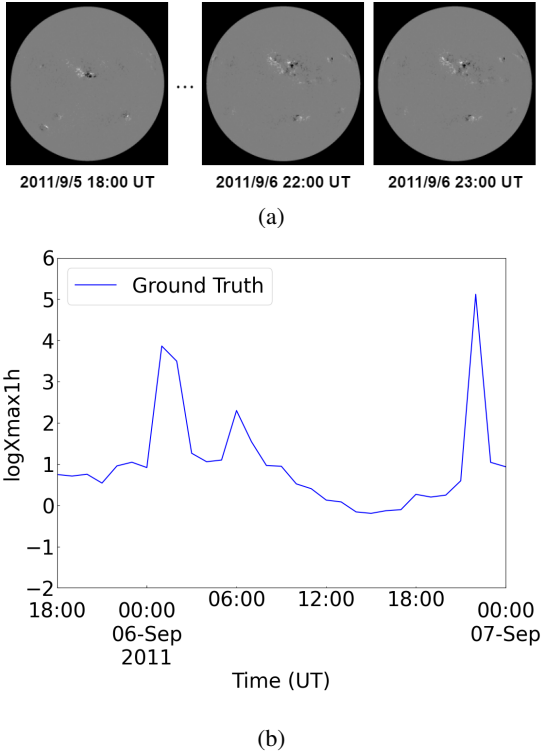


図1 対象タスクの例。(a) 磁場画像、(b) 太陽フレアのX線強度。

は確率的予測を行うよう最適化されたネットワーク構造となっている。さらに、Flare Transformer [25] では、物理特徴量に加えて磁場画像を使用し、Transformer 型注意機構による太陽フレア予測を行う。

太陽フレア予測において、標準的なデータセットは複数存在する。[11] は Solar Dynamic Observatory (SDO) [16] および Geostationary Operational Environmental Satellite (GOES) の撮影画像から抽出した物理特徴量を含むデータセットを提案している¹⁾

3 問題設定

本論文では、時刻 t から 24 時間後までにおける太陽フレアの X 線強度 $\mathbf{y}_{t+1:t+24} = \{y_{t+1}, y_{t+2}, \dots, y_{t+24}\}$ を予測するタスクを扱う。ここに、 y_t はそれぞれ時刻 t から 1 時間以内の最大 X 線強度を表す。

本タスクにおいて、以下の入出力を想定する。

- **入力:** 磁場画像および物理特徴量。
- **出力:** 時刻 t から 24 時間後までにおける太陽フレアの X 線強度の対数値。

本タスクでは、モデル出力 $\hat{\mathbf{y}}_{t+1:t+24}$ が $\mathbf{y}_{t+1:t+24}$ にできるだけ近いことが望ましい。

図1に代表例を示す。図1(a)は、2011年9月5日18:00 UTから2011年9月6日23:00 UTまでの磁場画像を表す。図1(b)において縦軸は太陽フレアのX線強度 ($\log X_{\max 1h}$)、横軸は時刻を表す。ここで縦軸の値は標準化を行っている。図中の青線は Ground Truth を表す。

本研究では、SDO/HMI [19] で撮影された太陽画像を磁場画像と定義する。タスクの評価尺度には、平均予測軌道誤差 $\bar{E}_{t+1:t+24}$ を用いる。

1) <https://wdc.nict.go.jp/IONO/wdc/solarflare/index.html>

4 提案手法

4.1 Flare Transformer Regressor

提案手法は Flare Transformer [25] を拡張したものである。Flare Transformer は太陽フレアクラスを予測するタスクにおいて良好な結果が報告されている。また、Flare Transformer は物理特徴量および太陽全体の磁場画像を入力として予測を行う。

ネットワーク入力 \mathbf{x} を以下のように定義する。

$$\mathbf{x} = (V_{t-k+1:t}, F_{t-k+1:t})$$

$$V_{t-k+1:t} = (v_{t-k+1}, v_{t-k+2}, \dots, v_t)$$

$$F_{t-k+1:t} = (f_{t-k+1}, f_{t-k+2}, \dots, f_t)$$

ここで、 $v_t \in \mathbb{R}^{512 \times 512}$ と $f_t \in \mathbb{R}^{90}$ はそれぞれ時刻 t における磁場画像および物理特徴量を示す。

ネットワークの構造を図2に示す。提案手法は主に3つのモジュールから構成され、それぞれ Magnetogram Module (MM), Sunspot Feature Module (SFM) および Transformer Decoder (TD) である。

4.2 Magnetogram Module

MMにおける入力は磁場画像 $V_{t-k+1:t}$ であり、 $V_{t-k+1:t}$ を以下のように計算する。

$$h_V = f_{FE}(V_{t-k+1:t})$$

ここに、 f_{FE} は Image Feature Extractor を表す。Image Feature Extractor は Vision Transformer [3] によって構成されている。Vision Transformer は Masked Autoencoder [6] を用いて事前学習を行っており、重みは固定されている。

Vision Transformer [3] の事前学習として、Masked Autoencoder [6] に基づき、入力に磁場画像 v を用いた画像の再構成タスクを行った。Masked Autoencoder は MAE encoder および MAE decoder の2つのモジュールから構成される。

まず、 v を 8×8 のパッチに分割し、 $P = \{p_1, p_2, \dots, p_{1024}\}$ を得る。ここに p_i は i 番目のパッチを表す。 p_i 内の全画素の標準偏差を計算し、 P のうち、標準偏差の小さい50%を P_{mask} 、それ以外のパッチを P_{enc} とする。MAE encoder の入力には P_{enc} のみを使用し、 P_{mask} は除去する。[6]では、全体の75%のパッチを除去していた。このとき、除去するパッチに重要領域が含まれると、特徴抽出の際にスパースな重要領域の情報欠如する。しかし、重要領域はパッチサイズより小さいため、周囲の情報から重要領域を再構成することは極めて困難である。そこで、スパースな重要領域が除去されないようにパッチ内の全画素の標準偏差をもとに除去するパッチを選択するという方法を採用する。

MAE encoder では、Vision Transformer [3] を用いて、 P_{enc} から画像特徴量 h_{enc} を抽出する。MAE decoder では、MAE encoder の出力 h_{enc} と、マスクトークン m を結合した $\{h_{\text{enc}}; m\}$ を入力とし、再構成された画像 \hat{v} を出力する。ここに m は予測すべき除去されたパッチ P_{mask} の存在を表す、学習可能ベクトルを表す。損失関数には、 v と \hat{v} の平均二乗誤差 (MSE) を計算する。ただし、損失はマスクされたパッチに対してのみ計算される。図3に磁場画像での再構成結果の例を示す。図は左から順に、あ

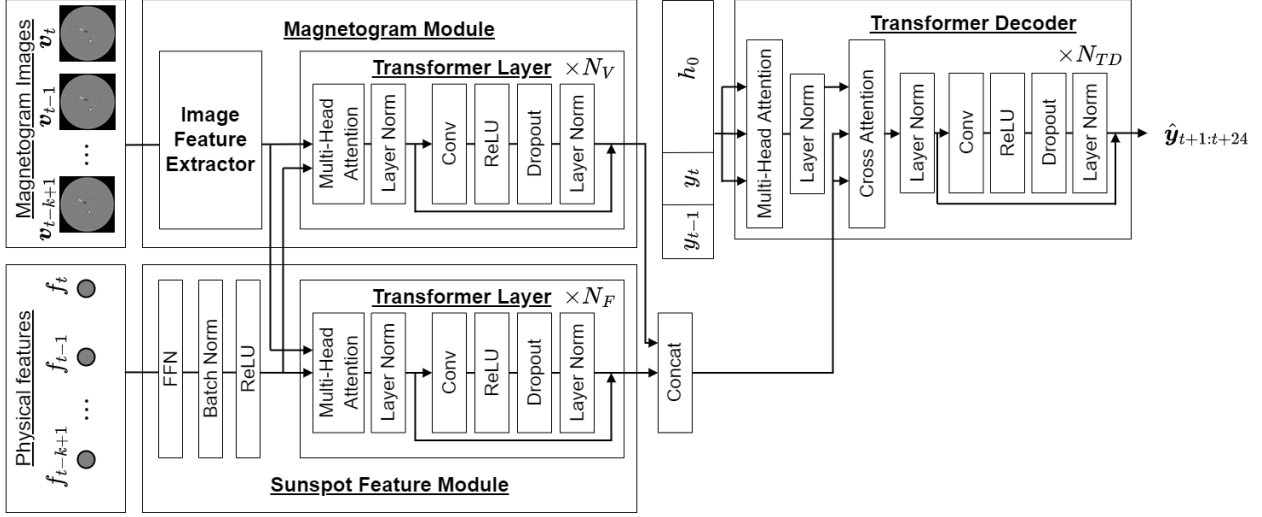


図2 提案手法のネットワーク図。

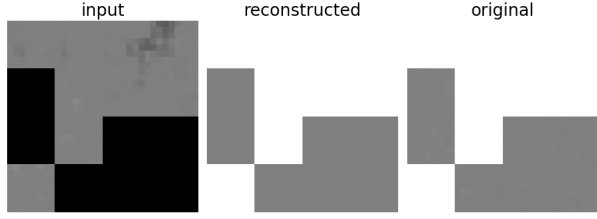


図3 Masked Autoencoder による再構成結果の例。

る磁場画像について同一領域における P_{enc} , \hat{v} , Ground Truth を表す。ただし、左図の黒く塗りつぶされた領域は除外されたパッチを表し、MAE encoder の入力には含まれない。また、中央図、右図について、損失に含まれないパッチは白色に塗りつぶされている。

次に、 N_V 層の transformer 層において、 h_V および h_{VF} をキー・バリューとして磁場画像および物理特徴量の時系列関係が計算される。

まず、Multi-head attention 層にて、 h_V および後述の h_{VF} がそれぞれ $h_V^{(i)} \in \mathbb{R}^{k \times d}$ および $h_{VF}^{(i)} \in \mathbb{R}^{k \times 2d}$ ($i = 1, \dots, N_{head}$) に分割される。ここに、 $d = H/N_{head}$ であり、 H は隠れ層の数、 N_{head} はヘッド数を表す。

次に、クエリ $Q^{(i)} \in \mathbb{R}^{k \times d}$ 、キー $K^{(i)} \in \mathbb{R}^{k \times 2d}$ としてバリュー $V^{(i)} \in \mathbb{R}^{k \times 2d}$ が以下のように計算される。

$$Q^{(i)} = W_q^{(i)} h_V^{(i)}, K^{(i)} = W_k^{(i)} h_{VF}^{(i)}, V^{(i)} = W_v^{(i)} h_{VF}^{(i)}$$

ここに、 $W_q^{(i)}$, $W_k^{(i)}$, $W_v^{(i)}$ はそれぞれ $Q^{(i)}$, $K^{(i)}$, $V^{(i)}$ の重みを表す。transformer 層の出力 h_{trm} は以下のように計算される。

$$\begin{aligned} h_{trm} &= f_{BN}(f_{FFN}(h_{mha})) \\ h_{mha} &= \left(f_{attn}^{(1)}, \dots, f_{attn}^{(N_{head})} \right) \\ f_{attn}^{(i)} &= \text{softmax} \left(\frac{Q^{(i)} K^{(i)T}}{\sqrt{d}} \right) V^{(i)} \end{aligned}$$

最終的な MM の出力 h_{MM} は N_V 層の transformer 層によって計算される。

4.3 Sunspot Feature Module

MM と同様に、SFM において物理特徴量 $F_{t-k+1:t}$ を以下のように計算する。

$$h_F = f_{BN}(f_{FFN}(F_{t-k+1:t}))$$

ここに、 f_{BN} は正規化層、 f_{FFN} は FFN を表す。その後、 h_V と h_F を結合することで h_{VF} を得る。

次に、MM と同様に N_F 層の transformer 層による処理を行い、SFM の出力 h_{SFM} を得る。

4.4 Transformer Decoder

TD における入力は h_{MM} , h_{SFM} および以下の式で表される h_{TD} である。

$$h_{TD} = \{y_{t-1}, y_t, h_0\}$$

ここに、 h_0 は目標系列長と同じ長さのゼロ埋めされた系列を表す。

TD では、まず h_{MM} , h_{SFM} が結合され、 $h_{MM;SFM}$ を得る。次に、 N_{TD} 層の transformer 層において、MM, SFM と同様の処理を行う。ただし、クエリを h_{TD} を基に作成し、キーおよびバリューを $h_{MM;SFM}$ を基に作成するという違いがある。最終的に TD から $\hat{y}_{t+1:t+24}$ を得る。損失関数 \mathcal{L} は以下のように定義される。

$$\mathcal{L} = \frac{1}{24} \sum_{k=1}^{24} (y_{t+k} - \hat{y}_{t+k})^2$$

5 実験

5.1 データセットおよび実験設定

実験では、SDO のウェブアーカイブ¹⁾および物理特徴量データベース²⁾から収集したデータセットを用いて本手法を評価した。本データセットには、2010年6月から2017年12月における61315のサンプルが含まれている。

本データセットに対して、次のように事前処理を行った。まず、時刻 t において複数の活動領域が検出されて

1) <https://sdo.gsfc.nasa.gov/data/>

2) <https://wdc.nict.go.jp/IONO/wdc/solarflare/index.html>

表1 各手法における定量的結果.

	$\bar{E}_{t+1:t+24}$
Flare Transformer [25]	1.06 ± 0.44
(i) Ours(ViT [3]→CNN)	0.52 ± 0.10
(ii) Ours(Full Attention→ProbSparse Attention)	0.50 ± 0.02
(iii) Ours	0.48 ± 0.03

いた場合、 y_t が最大である、黒点に関する物理特徴量を選択した。次に、時刻 t において活動領域が検出されなかった場合、 t より前の時刻において最後に観測された黒点に関する物理特徴量および y を選択した。

本研究では時系列予測において標準的な手法である時系列交差検証 [21] に基づき、訓練集合、検証集合およびテスト集合を分割した。ここで時系列交差検証では、訓練サンプルは検証集合およびテスト集合を構成するサンプルよりも時系列的に前のサンプルで構成される。

訓練集合、検証集合およびテスト集合はそれぞれパラメータの学習、ハイパーパラメータの決定、性能の評価に使用した。ハイパーパラメータは、 N_V を 2、 N_F を 2 とした。最適化には Adam を使用し、学習率は 1×10^{-6} 、バッチサイズは 8 とした。提案手法の訓練可能パラメータ数、および積和演算数は、それぞれ 280 万、および 6600 万である。学習にはメモリ 11GB 搭載の GeForce RTX 2080 および Intel Core i9-9900K を使用した。本データセットにおけるモデルの訓練時間および 1 サンプルあたりの推論時間は、それぞれ 3 時間および 5.0×10^{-3} 秒であった。また、検証集合における損失関数の値が 3 回連続改善しなかった場合に early stopping を行った。

5.2 実験結果

提案手法の有効性を検証するため、ベースライン手法との比較実験を行った。表 1 に定量的結果を示す。表には各実験における各年度ごとの平均値および標準偏差を示した。ベースライン手法は、Flare Transformer [25] とした。Flare Transformer は太陽フレアクラスを予測するタスクにおいて良好な結果が報告されており、また、物理特徴量と磁場画像の両方を入力として扱う手法であるため、本モデルのベースライン手法とする。

タスクの評価尺度には、平均予測軌道誤差 $\bar{E}_{t+1:t+24}$ を用いる。ここで $\bar{E}_{t+1:t+24}$ を以下のように定義する。

$$\bar{E}_{t+1:t+24} = \frac{1}{N} \sum_{i=1}^N E_{t+1:t+24}^{(i)}$$

$$E_{t+1:t+24}^{(i)} = \frac{1}{24} \sum_{k=1}^{24} \left(y_{t+k}^{(i)} - \hat{y}_{t+k}^{(i)} \right)^2$$

ここに、 i 、 N および $E_{t+1:t+24}^{(i)}$ はそれぞれサンプルのインデックス、サンプル数および 24 時間の予測軌道誤差を表す。 $\bar{E}_{t+1:t+24}$ は本タスクにおいて標準的な尺度であるため、使用した。

表 1 より、 $\bar{E}_{t+1:t+24}$ において提案手法とベースライン手法はそれぞれ、 0.48 ± 0.03 および 1.06 ± 0.44 であり、提案手法が 0.58 ポイント改善した。したがって、ベースライン手法と比較して提案手法が優れているという結果を得た。

図 4 に定性的結果を示す。図 4 は提案手法が成功した例を表す。図において縦軸は太陽フレアの X 線強度、横

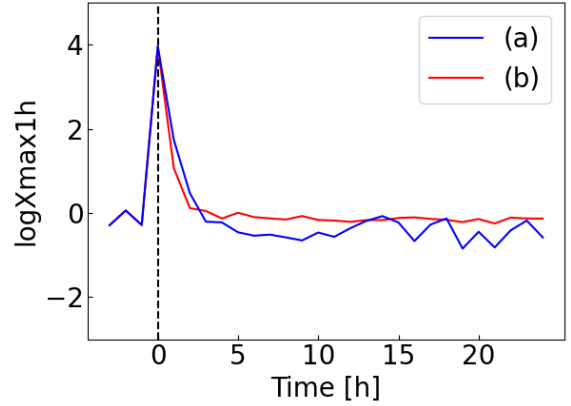


図4 定性的結果 (成功例). (a)Ground Truth, (b) 提案手法.

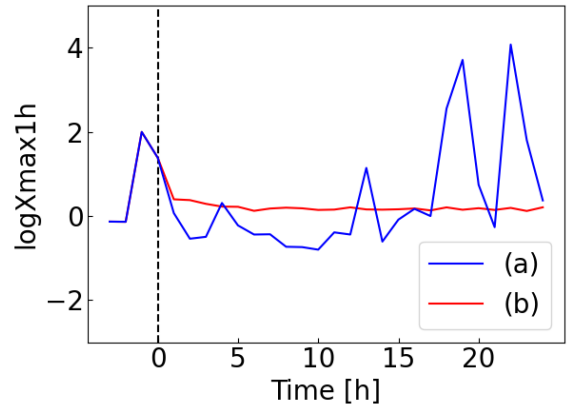


図5 定性的結果 (失敗例). (a)Ground Truth, (b) 提案手法.

軸は時間を表す。ここで縦軸の値は標準化を行っている。図中の青線および赤線はそれぞれ Ground Truth および予測結果を表す。また、破線は予測開始時刻を示し、破線より左側は入力、右側は予測結果を表す。図 4 において予測開始時刻は、2016 年 4 月 18 日 00:00 UT である。図 4 では、入力である y_{t-3} から y_{t-1} は概ね増加する時系列であるが、提案手法は y_{t+1} から y_{t+10} の減少傾向を捉えることに成功している。

図 5 に、提案手法が失敗した例を示す。予測開始時刻は 2016 年 7 月 22 日 7:00 UT である。この例では、時刻 $t+18$ や時刻 $t+22$ において 4 程度の大きな誤差があり、予測に失敗していることがわかる。失敗した原因は、 y_{t-1} までは減少傾向にあったが、時刻 $t+16$ から $t+25$ にかけて y_t が急激な増減を繰り返していたためだと考えられる。

5.3 Ablation studies

Ablation studies として、以下の条件を定めた。

- (i) 本モデルにおける Vision Transformer [3] で構成される Image Feature Extractor を、Frare Transformer [25] における CNN で構成される Image Feature Extractor へと置き換えた場合、性能にどの程度の差が生じるかを調べた。

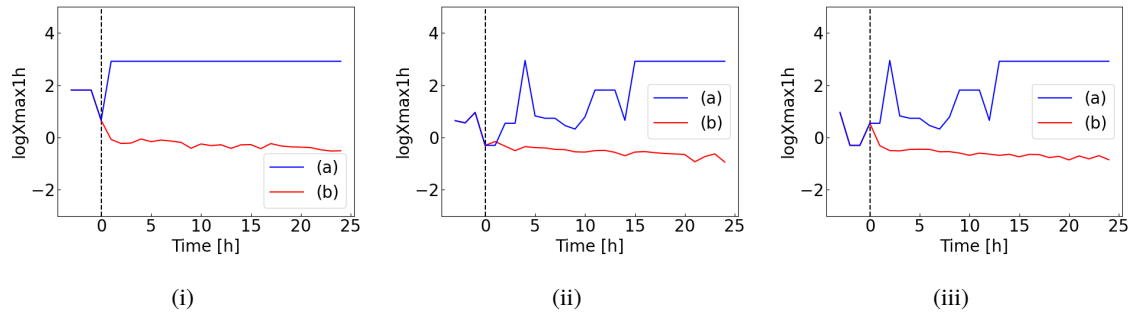


図6 失敗要因の例. (i) 欠損値に関する誤り, (ii) $y_{t+1:t+24}$ が $y_{t-3:t}$ と傾向が異なることによる失敗例, (iii) 深刻な誤り.

表2 失敗例の分類.

エラー ID	詳細	#Errors
(A)	欠損値に関する誤り	27
(B)	$y_{t+1:t+24}$ が $y_{t-3:t}$ と傾向が異なる	2
(C)	時系列予測に関する誤り	1

(ii)本モデルで使われている Multi-head Attention を Informer [24] で用いられる ProbSparse Attention に置き換えた場合, 性能にどの程度の差が生じるかを調べた.

Probsparse Attention は, 以下で表される sparsity measurement M のもと, 上位 u 個のクエリのみを attention の計算に使用する. これにより, 系列長 L に対して計算量 $O(L \log L)$ で計算できる.

$$M(q_i, K) = \max_j \frac{q_i k_j^\top}{\sqrt{d}} - \frac{1}{L_K} \sum_{j=1}^{L_K} \frac{q_i k_j^\top}{\sqrt{d}}$$

ここに, q_i, k_i はそれぞれ Q, K の i 行目を表す.

表1より, 条件 (i) では, $\bar{E}_{t+1:t+24}$ が 0.04 ポイント悪化したことから, Vision Transformer [3] で構成される Image Feature Extractor を使用したことがモデルの性能向上に寄与していると考えられる. また, 条件 (ii) では $\bar{E}_{t+1:t+24}$ が 0.04 ポイント悪化した, これは今回用いたモデルでは $k = 4$ と小さく, ProbSparse Attention において, sparsity measurement M のもと上位 u 個のクエリのみを Attention の計算に使用したことが, かえって予測性能に悪影響を与えたためであると考えられる. k を大きくしたときの性能への影響については, 今後の検討課題とする.

5.4 エラー分析

最後に, エラー分析を行う. 失敗例の分類を表2に示す. $E_{t+1:t+24}$ が 0.8 より大きい 487 サンプルのうち無作為に抽出した 30 サンプルの原因を以下に示すように3つに大別する.

(A)欠損値に関する誤り

本研究では, データセットにおける欠損値を, それ以前の時刻において最後に観測された値で置き換えた. (A) は, この部分について誤差が大きい場合を表す. 図6(i)に(A)の例を示す. この例では, 時刻 $t+2$ から時刻 $t+24$ において y_t が欠損しているため, y_{t+2} から y_{t+24} は y_{t+1} を用いている. この例では, 値が欠損している時刻 $t+2$ から時刻 $t+24$ において誤差が大きい.

(B) $y_{t+1:t+24}$ が $y_{t-3:t}$ と傾向が異なる

入力系列 $y_{t-3:t}$ に現れない傾向が, 予測対象系列 $y_{t+1:t+24}$ に現れている場合を表す. 図6(ii)に(B)の例を示す. この例における失敗の原因は, y_{t-1} までは減少傾向にあったが, 時刻 $t+1$ から $t+16$ にかけて y_t が増減を繰り返していたことで, 予測が困難であったためだと考えられる.

(C)深刻な誤り

提案手法が過去の情報を考慮することに失敗している場合を表す. 図6(iii)に(C)の例を示す. この例では, y_{t-1} から y_t が増加傾向にあるにもかかわらず, 時刻 $t+1$ 以降も増加する y_t の予測に失敗している.

表2より, 主要なエラー要因は欠損値部分の誤差が大きい点であるといえる. これに対して, 欠損値部分を評価から取り除くといった方法が考えられる.

6 おわりに

本論文では, 時刻 t から 24 時間後までの太陽フレアの X 線強度 $y_{t+1:t+24}$ を予測するタスクを扱った. 本研究の貢献を以下に示す.

- 太陽フレア予測に寄与する特徴量を抽出するため, Masked Autoencoder [6] で事前学習した ViT [3] を導入した.
- 太陽フレア予測を回帰問題として扱うために, Informer [24] を参考にした回帰用デコーダを構築した.
- 標準データセット上で, 提案手法はベースライン手法を上回る性能を達成した.

将来研究として, 24 時間後以降におけるフレア予測が考えられる.

謝辞

本研究の一部は, JSPS 科研費 20H04269, NEDO の助成を受けて実施されたものである.

参考文献

- [1] Shamik Bhattacharjee, Rasha Alshehhi, Dattaraj Dhuri, Rasha Alshehhi, Dattaraj Dhuri et al., “Supervised Convolutional Neural Networks for Classification of Flaring and Nonflaring Active Regions using Line-of-sight Magnetograms”, *The Astrophysical Journal*, Vol. 898, No. 2, p. 98 (12pp) (2020)
- [2] George Box, Gwilym Jenkins, Gregory Reinsel et al., *Time Series Analysis: Forecasting and Control*, John Wiley & Sons (2015)
- [3] Alexey Dosovitskiy, Lucas Beyer et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”, in *ICLR* (2021)
- [4] Philippe Esling et al., “Time-series data mining”, *ACM Computing*

- Surveys*, Vol. 45, No. 1, pp. 1–34 (2012)
- [5] Manolis Georgoulis , Shaun Bloomfield et al., “The flare likelihood and region eruption forecasting (FLARECAST) project: flare forecasting in the big data & machine learning era”, *Journal of Space Weather and Space Climate*, Vol. 11, p. A39 (37pp) (2021)
- [6] Kaiming He , Xinlei Chen , Saining Xie , Yanghao Li , Piotr Dollár et al., “Masked autoencoders are scalable vision learners”, in *CVPR*, pp. 16000–16009 (2022)
- [7] Rob Hyndman , Anne Koehler et al., *Forecasting with Exponential Smoothing: The State Space Approach*, Springer Science & Business Media (2008)
- [8] Yūki Kubo , Mitsue Den et al., “Verification of Operational Solar Flare Forecast: Case of Regional Warning Center Japan”, *Journal of Space Weather and Space Climate*, Vol. 7, p. A20 (16pp) (2017)
- [9] Kanya Kusano , Tomoya Iju et al., “A Physics-based Method that can Predict Imminent Large Solar Flares”, *Science*, Vol. 369, No. 6503, pp. 587–591 (2020)
- [10] Shiyang Li et al., “Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting”, in *NeurIPS*, Vol. 32, pp. 5243–5253 (2019)
- [11] Naoto Nishizuka et al., “Solar flare prediction model with three machine-learning algorithms using ultraviolet brightening and vector magnetograms”, *The Astrophysical Journal*, Vol. 835, No. 2, p. 156 (2017)
- [12] Naoto Nishizuka , Komei Sugiura et al., “Deep flare net (DeFN) model for solar flare prediction”, *The Astrophysical Journal*, Vol. 858, No. 2, p. 113 (8pp) (2018)
- [13] Naoto Nishizuka , Yuki Kubo , Komei Sugiura , Mitsue Den et al., “Reliable Probability Forecast of Solar Flares: Deep Flare Net-Reliable (DeFN-R)”, *The Astrophysical Journal*, Vol. 899, No. 2, p. 150 (8pp) (2020)
- [14] Bun Theang Ong et al., “Dynamically Pre-trained Deep Recurrent Neural Networks using Environmental Monitoring Data for Predicting PM2.5”, *Neural Computing and Applications*, Vol. 27, No. 6, pp. 1553–1566 (2016)
- [15] Eunsu Park , Yong-Jae Moon , Seulki Shin et al., “Application of the Deep Convolutional Neural Network to the Forecast of Solar Flare Occurrence Using Full-disk Solar Magnetograms”, *The Astrophysical Journal*, Vol. 869, No. 2, p. 91 (6pp) (2018)
- [16] William Pesnell , Barbara Thompson et al., “The Solar Dynamics Observatory (SDO)”, in *The Solar Dynamics Observatory*, pp. 3–15, Springer (2011)
- [17] Syama Rangapuram , Matthias Seeger et al., “Deep State Space Models for Time Series Forecasting”, in *NeurIPS*, Vol. 31, pp. 7785–7794 (2018)
- [18] David Salinas , Valentin Flunkert , Jan Gasthaus et al., “DeepAR: Probabilistic forecasting with autoregressive recurrent networks”, *International Journal of Forecasting*, Vol. 36, No. 3, pp. 1181–1191 (2020)
- [19] Philip Scherrer , Jesper Schou , R Bush , A Kosovichev , R Bogart et al., “The Helioseismic and Magnetic Imager (HMI) Investigation for the Solar Dynamics Observatory (SDO)”, *Solar Physics*, Vol. 275, No. 1, pp. 207–227 (2012)
- [20] SwissRe , “Solar storm; how to calculate insured / reinsured losses?”, Space Weather Workshop (2016)
- [21] Leonard Tashman , “Out-of-sample tests of forecasting accuracy: an analysis and review”, *International Journal of Forecasting*, Vol. 16, No. 4, pp. 437–450 (2000)
- [22] Ruofeng Wen , Kari Torkkola et al., “A Multi-horizon Quantile Recurrent Forecaster”, *arXiv preprint arXiv:1711.11053* (2017)
- [23] Sifan Wu , Xi Xiao , Qianggang Ding , Peilin Zhao et al., “Adversarial Sparse Transformer for Time Series Forecasting”, in *NeurIPS*, Vol. 33, pp. 17105–17115 (2020)
- [24] Haoyi Zhou , Shanghang Zhang , Jieqi Peng , Shuai Zhang , Jianxin Li et al., “Informer: Beyond Efficient Transformer for Long Sequence Time-series Forecasting”, in *AAAI*, pp. 11106–11115 (2021)
- [25] 兼田 寛大, 飯田 紡 et al., “Flare transformer: 磁場画像と物理特徴量を用いた太陽フレア予測”, 人工知能学会全国大会論文集 第36回全国大会 (2022) 一般社団法人人工知能学会 (2022)