

Flareformer: 磁場画像および物理特徴量の統合による大規模太陽フレア予測 Flareformer: Large-scale Solar Flare Prediction by Integrating Magnetograms and Sunspot Physical Features

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

1 はじめに

太陽フレアに代表される大規模な太陽活動は、電子機器の障害や宇宙飛行士への健康的被害など様々な影響をもたらす。実際、経済損失に関する試算では、Carrington 級フレアの発生で約 1,630 億ドルの損失が生じると推定されており [20]、太陽フレアの予測は極めて重要なタスクであるといえる。

しかし、24 時間以内に発生する最大の太陽フレアクラスを予測するタスクは容易ではない。例えば、完璧な予測に対して 1.0 を返す BSS(Brier Skill Score) [13] や GMGS(Gandin-Murphy-Gerrity score) [2] の指標において、2010 年から 2015 年における専門家予測は $BSS_{\geq M} = 0.16$ および $GMGS = 0.48$ 程度である [5, 10]。

こうした背景から、本論文では、時刻 t から 24 時間以内に発生する最大の太陽フレアクラスを予測するタスクを扱う。本タスクにおける入力は、時刻 $(t-k, t]$ における一時間おきの太陽の磁場画像と物理特徴量であり、出力は 24 時間以内における最大の太陽フレアクラスの発生確率である。ここで、物理特徴量とは AIA(Atmospheric Imaging Assembly) [7] や SDO(Solar Dynamic Observatory) [16] による撮影画像から抽出した 90 種類の特徴量 [11] である。また太陽フレアクラスは X 線ピーク値 p_t により定義され、 p_t が低い順に O, C, M, X の 4 クラスが存在する。

既存手法では、太陽フレアクラスを実際よりも低く見積る予測が多い。例えば Flare Transformer [25] では、X クラスの太陽フレアを高頻度で M クラス以下と予測してしまう。したがって本論文では、クラスバランスの不均衡性を考慮した、正確で信頼性の高い太陽フレア予測モデル Flareformer を提案する。

提案手法における新規性は以下の通りである。

- 太陽フレア予測を扱う提案手法において、時系列予測に適した構造である Informer [24] の注意機構を導入する。
- 太陽フレア予測に寄与する特徴量を抽出するため、ConvNeXt [9] を導入する。
- サンプル数を伴う訓練時において、サンプル数の少ない X クラスへの過学習を防ぐため、Classifier Re-training (cRT) [4] に基づく分類器の再学習機構を導入する。

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]。時系列予測分野では、統計モデルに基づく古典的手法 [1, 3] のほか、近年には DNN(Deep Neural Networks) を用いた手法も数多く提案されている。DNN による時系列予測モデルは二つに大別でき、RNN(Recurrent Neural Network) ベースの手法 [14, 17, 18, 22]、および Attention ベースの手法 [8, 23, 24] がある。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 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。

太陽フレア予測の研究も盛んに行われている [6, 12, 15]。DeFN [12] は 24 時間以内における太陽フレアの発生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。DeFN を拡張した DeFN-R [13] では予測の信頼性に適し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。

3 問題設定

本論文では、時刻 t から 24 時間以内に発生する最大の太陽フレアクラスを予測するタスクを扱う。ここで、太陽フレアクラスとは X 線ピーク値による太陽フレアの等級を指し、時刻 t における太陽フレアクラスは X 線ピーク値 p_t により以下のように定義される。

$$y_t = \text{flareclass}(\max\{p_{t+1}, p_{t+2}, \dots, p_{t+24}\})$$

本論文では、以下の入出力を想定する。

- **入力:** 時刻 $(t-k, t]$ における一時間ごとの磁場画像 $V_{t-k+1:t}$ および物理特徴量 $F_{t-k+1:t}$ 。
- **出力:** 各太陽フレアクラスに対する予測確率を示す 4 次元ベクトル $p(\hat{y}_t)$ 。

上記タスクは太陽フレア予測の分野において標準的な問題設定であり [12, 15]、専門家による予測も同様に分類問題として取り組まれている。したがって、本論文では回帰問題ではなく上述の分類問題を扱う。

本タスクでは、モデル出力 $p(\hat{y}_t)$ と y_t が等しくなることが望ましい。しかし、太陽フレアは各クラスの発生頻度が不均衡であるため、各クラスを均等に評価する損失関数を用いた場合、全体の 90% を占める O クラスや C クラスといった自明な解を出力する。こうした自明な解を

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

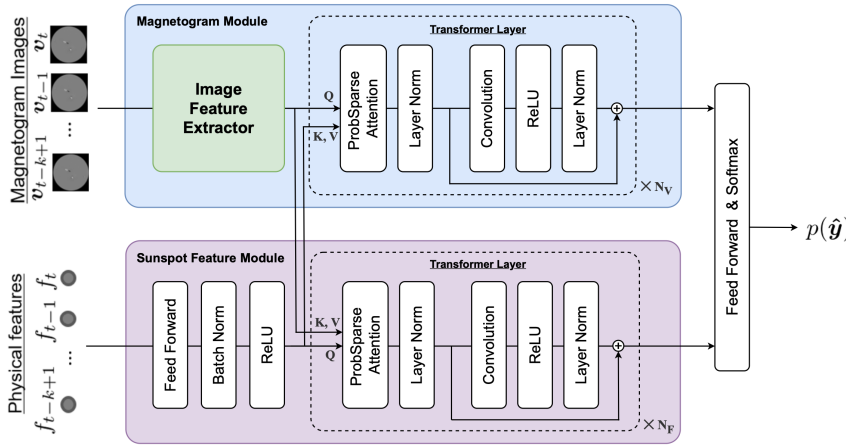


図1 提案手法のネットワーク構造

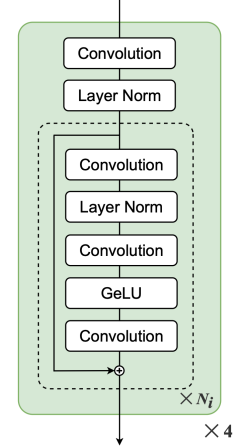


図2 Image Feature Extractor の構造

避けるため、本タスクでは、本分野において標準的な尺度である BSS [13] および GMGS [2] を用いて評価する。したがって、本タスクにおける評価尺度には、 $BSS_{\geq M}$ 、GMGS [2]、および $TSS_{\geq M}$ を使用する。ここで、“ $\geq M$ ”は太陽フレアクラスが M クラス以上または M クラス未満の二値として評価したことを示す。

本論文では、太陽全体の画像から太陽フレアクラスを予測することを前提とする。そのため、黒点領域レベルにおける予測 [12] は扱わない。

4 提案手法

提案手法¹⁾のネットワーク構造を図1に示す。図における“Batch Norm”、“Layer Norm”はそれぞれバッチ正規化層、およびレイヤ正規化層を示す。

本手法は磁場画像と物理特徴量を入力とする Flare Transformer [25] を拡張したものであり、ネットワークは Magnetogram Module (MM) および Sunspot Feature Module (SFM) の2つのモジュールから構成される。各モジュールの詳細については本章にて後述する。

4.1 入力

時刻 $t-k+1$ から時刻 t における、ネットワークの入力 x を以下のように定義する。

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

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

$$F_{t-k+1:t} = (f_{t-k+1}, f_{t-k+2}, \dots, f_t) \quad (3)$$

ここで、 $v_t \in \mathbb{R}^{256 \times 256}$ 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。

4.2 Magnetogram Module

Magnetogram Module では、磁場画像 $V_{t-k+1:t}$ を用いて h_V が以下のように計算される。

$$h_V = f_{FE}(V_{t-k+1:t}) \quad (4)$$

ここで、 f_{FE} は図2に示す Image Feature Extractor を指す。図中 N_i は i stage 目での層数を表しており、各 stage において $\{N_i\}_{i=0}^4 = \{3, 3, 9, 3\}$ である。すなわち、Image Feature Extractor は 18 層の ConvNeXt [9] で構成される。

4.3 Sunspot Feature Module

Sunspot Feature Module では、物理特徴量 $F_{t-k+1:t}$ を用いて h_F が以下のように計算される。

$$h_F = \text{ReLU}(f_{BN}(f_{FFN}(F_{t-k+1:t}))) \quad (5)$$

ここで、 f_{BN} はバッチ正規化層、 f_{FFN} は FFN(Feed Forward Network) を示す。上式による計算後、 h_V と h_F を結合することで h_{VF} を得る。

4.4 Transformer Layers

Transformer 層では、磁場画像および物理特徴量の時系列関係が計算される。Transformer 層に入力するクエリを q とすると、MM では $q = h_V$ であり、SFM では $q = h_F$ である。また、入力されるキーとバリューは MM、SFM 共に h_{VF} である。

ProbSparse Attention 層 [24] では、 q 、 h_{VF} がそれぞれ $q^{(i)} \in \mathbb{R}^{k \times d}$ および $h_{VF}^{(i)} \in \mathbb{R}^{k \times 2d}$ ($i = 1, \dots, N_{\text{head}}$) に分割される。ここで $d = H/N_{\text{head}}$ であり、 H は隠れ層の数、 N_{head} はヘッド数を示す。

クエリ $Q^{(i)} \in \mathbb{R}^{k \times d}$ 、キー $K^{(i)} \in \mathbb{R}^{k \times 2d}$ 、およびバリュー $V^{(i)} \in \mathbb{R}^{k \times 2d}$ が i 番目のヘッドについて以下のように計算される。

$$Q^{(i)} = W_q^{(i)} q^{(i)}, K^{(i)} = W_k^{(i)} h_{VF}^{(i)}, V^{(i)} = W_v^{(i)} h_{VF}^{(i)} \quad (6)$$

ここで、 $W_q^{(i)}$ 、 $W_k^{(i)}$ 、 $W_v^{(i)}$ 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。

$$h_{\text{trm}} = f_{\text{LN}}(h_{\text{mid}}) + h_{\text{mha}} \quad (7)$$

$$h_{\text{mid}} = \text{ReLU}((f_{\text{conv}}(h_{\text{mha}}))) \quad (8)$$

$$h_{\text{mha}} = f_{\text{LN}}\left(\left[f_{\text{attn}}^{(1)}; f_{\text{attn}}^{(2)}; \dots; f_{\text{attn}}^{(N_{\text{head}})}\right]\right) \quad (9)$$

$$f_{\text{attn}}^{(i)} = \text{softmax}\left(\frac{Q^{(i)} K^{(i)\top}}{\sqrt{d}}\right) V^{(i)} \quad (10)$$

ここで、 f_{LN} はレイヤ正規化層、 f_{conv} は畳み込み層を示す。

MM の出力 h_{MM} は N_V 層の transformer 層によって計算され、SFM の出力 h_{SFM} も同様に N_F 層の transformer 層によって計算される。

1) <https://github.com/keio-smilab22/Flareformer>

最後に、予測クラス y_t^* は以下のように計算される。

$$y_t^* = \operatorname{argmax}_i(p(\hat{y}_{ti})) \quad (11)$$

$$p(\hat{y}_t) = \operatorname{softmax}(f_{\text{FFN}}(\mathbf{h})) \quad (12)$$

$$\mathbf{h} = [\mathbf{h}_{\text{MM}}; \mathbf{h}_{\text{SFM}}] \quad (13)$$

ここで、 $p(\hat{y}_{ti})$ は i 番目のクラスに関する予測確率を示す。

4.5 Classifier Re-training (cRT)

本タスクはクラスバランスが偏っている。例えば、2017 年における X クラスのサンプル数は全体の約 2.9% であった。したがって、学習時にはクラス間のサンプル数が均等になるようサンプリングする必要がある。

しかし、そうしたオーバーサンプリングを伴う学習は、サンプル数が極端に少ない X クラスへの過学習を招く。そこで、本提案手法では元のデータセットでの学習に加え、Classifier Re-training (cRT) [4] に基づくモデルの再学習を行った。ここで、元のデータセットでの学習段階を first-stage とし、cRT に基づく学習段階を second-stage と呼ぶことにする。

second-stage では、クラス間のサンプル数が均等になるようサンプリングを行い、また最後の Linear 層以外の重みを固定して再学習を行う。すなわち、式 (12) の指す最終層の重みのみを、クラス数の均されたデータセットで再学習させる。

4.6 損失関数

提案手法では、Flare Transformer [25] 同様、GMGS 損失および BSS 損失を導入する。

GMGS 損失は以下のように定義される。

$$\mathcal{L}_{\text{GMGS}} = -\frac{1}{NI} \sum_{n=1}^N s_{i^*j^*} \sum_{i=1}^I y'_{ni} \log(p(\hat{y}_{ni})) \quad (14)$$

$$i^* = \operatorname{argmax}_i(y_{ni}), j^* = \operatorname{argmax}_j(p(\hat{y}_{nj})) \quad (15)$$

ここで、 N 、 I および $s_{i^*j^*}$ はそれぞれサンプル数、クラス数、および GMGS のスコア行列 [2] を示す。また、 $p(\hat{y}_{ni})$ 、 y_{ni} 、そして y'_{ni} はそれぞれ n 番目のサンプルに対する i 番目のクラスの予測確率、正解ラベルを示す。

BSS 損失は以下のように定義される。

$$\mathcal{L}_{\text{BSS}} = -\frac{1}{NI} \sum_{n=1}^N \sum_{i=1}^I (p(\hat{y}_{ni}) - y_{ni})^2 \quad (16)$$

上記の GMGS 損失および BSS 損失を用いて、全体の損失関数を以下のように定義する。

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda_{\text{GMGS}} \mathcal{L}_{\text{GMGS}} + \lambda_{\text{BSS}} \mathcal{L}_{\text{BSS}} \quad (17)$$

ここで、 $\mathcal{L}_{\text{CE}}$ は $\mathbf{y}_n$ と $\hat{\mathbf{y}}_n$ に関する交差エントロピー損失を示し、 λ_{GMGS} および λ_{BSS} は重みを示す。

5 実験

5.1 実験設定

表 1 に実験設定を示す。 N_V 、 $\#H_V$ 、 $\#A_V$ はそれぞれ Magnetogram Module における transformer の層数、隠れ層の次元数、および attention のヘッド数を示す。また

表 1 提案手法のパラメータ設定

Magnetogram Module	$N_V = 1, \#H_V = 256$ $\#A_V = 4$
Sunspot Feature Module	$N_F = 1, \#H_F = 256$ $\#A_F = 4$
Loss weights	$\lambda_{\text{GMGS}} = 0.01, \lambda_{\text{BSS}} = 10$
Window size	$k = 4$
Optimizer	Adam ($\beta_1 = 0.9, \beta_2 = 0.999$)
Max learning rate	2.1×10^{-6}
Batch size	48

表 2 各手法における定量的結果

Method	BSS $_{\geq M} \uparrow$	GMGS $\uparrow$	TSS $_{\geq M} \uparrow$
DeFN [12]	-0.022 $\pm$ 0.78	0.375 $\pm$ 0.14	0.413 $\pm$ 0.15
DeFN-R [13]	0.036 $\pm$ 0.98	0.302 $\pm$ 0.055	0.279 $\pm$ 0.16
Flare Transformer [25]	0.082 $\pm$ 0.97	0.503 $\pm$ 0.059	0.530 $\pm$ 0.11
提案手法	0.22$\pm$0.68	0.500$\pm$0.19	0.530$\pm$0.11
Human [5] [10]	0.16	0.48	0.50

N_F 、 $\#H_F$ 、 $\#A_F$ は同様に Sunspot Feature Module における transformer のハイパーパラメータを示す。

本実験では、first-stage にて 40 epoch、second-stage にて 3 epoch 学習させた。また、first-stage では学習率のスケジューリングを行い、初期の 5epoch 間では学習率を線形に増加させ、cosine-decay によって学習率を減少させた。また second-stage では、再度学習率を元に戻し、再学習させた。

本モデルにおける訓練可能パラメータ数、および積和演算数は、それぞれ 6.7M、および 30.8G である。学習にはメモリ 16GB 搭載の GeForce RTX 3080 および Intel Core i9-11980HK を使用し、本モデルの訓練時間と 1 サンプルあたりの推論時間は、それぞれ 4.8 時間、 3.1×10^{-3} 秒であった。

5.2 データセット

本データセットは SDO のウェブアーカイブ²⁾ および物理特微量データベース³⁾ から収集した磁場画像および物理特微量によって構成されている。

本データセットには、2010 年 6 月 1 日 2 時から 2017 年 12 月 24 日 12 時までの合計 61315 枚の HMI [19] によって撮影された磁場画像、および各時刻における 90 種類の物理特微量 [12] が含まれる。また本データセットにおいて、正解ラベルが X、M、C、O クラスとなるサンプルの数はそれぞれ 492、4745、19736、36342 であった。

本論文では、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、訓練集合と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ている。

5.3 実験結果

本研究では、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。ただし、ベースライン手法を、DeFN [12]、DeFN-R [13]、および Flare Transformer [25] に設定した。

表 2 にベースライン手法と提案手法の定量的結果、お

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

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

よび専門家による予測評価を示す。表には各実験における年度ごとの平均値と標準偏差を示した。ただし、2016 年は X クラスの太陽フレアが観測されなかったため、GMGS については近似値を用いた。

本タスクにおける評価尺度には、 $BSS_{\geq M}$ 、GMGS [2]、および $TSS_{\geq M}$ を使用する。ここで、 BSS や GMGS は太陽フレア予測タスクにおいて標準的な尺度であり、本研究における主要尺度は $BSS_{\geq M}$ 、GMGS とする。

ベースライン手法の中で最も結果の良い手法は Flare Transformer [25] であった。そのため提案手法と Flare Transformer [25] の結果を比較する。表 2 より、主要尺度である $BSS_{\geq M}$ について、提案手法と Flare Transformer [25] はそれぞれ 0.082、および 0.222 であり、提案手法が 0.14 ポイント上回った。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。

また、表 2 に示した専門家予測の値と提案手法の結果を比較すると、 $BSS_{\geq M}$ 、GMGS、 $TSS_{\geq M}$ の値をそれぞれ、0.062, 0.03, 0.02 ポイント上回っており、提案手法は専門家を越える予測性能を達成することができた。

図 3 に定性的結果を示す。図 3(a) は、X クラスを正しく予測した例を示している。この例は、2017 年 9 月 5 日 6:00 UT から 9:00 UT における磁場画像のサンプルである。このサンプルにおける予測は $y_t^* = "X"$ であり、実際に 24 時間後の 2017 年 9 月 6 日 9:00 UT において X クラスの太陽フレアが発生したため、正しく太陽フレアクラスを予測できた。同様に、図 3(b) に M クラスを正しく予測した例を示す。

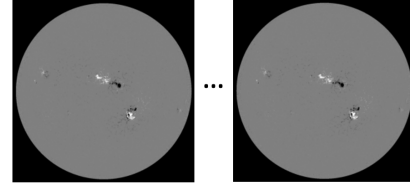
図 3(c) は正しいクラスを予測できなかった例を示している。この例では、2017 年 1 月 21 日 0:00 UT から 3:00 UT におけるサンプルを示しており、正解は $y_t = "C"$ である。本手法のモデルはこのサンプルについて $y_t^* = "M"$ と出力したため、予測に失敗した。ここで、C および M クラスの予測確率はそれぞれ、 $p(y_{t1}) = 0.391$, $p(y_{t2}) = 0.399$ と互いに近い値であるため、原因はこのサンプルが C クラスと M クラスの決定境界付近に位置しているためであると考えられる。

5.4 Ablation studies

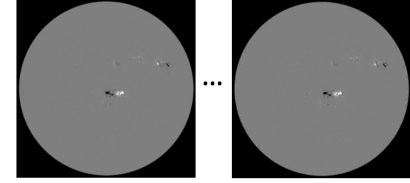
以下に示す 4 条件を ablation studies に定めた。

- (i) cRT による性能への寄与を調べるため、cRT を用いずに学習させた。
- (ii) Transformer 層の層数を増やした場合の性能を調べるため、層数 (N_V, N_F) を $(N_V, N_F) = (2, 2), (3, 3)$ と増やした。
- (iii) ConvNeXt [9] の性能への寄与を調べるため、画像特徴量を抽出する機構を Flare Transformer の Image Feature Extractor へと置き換えた。
- (iv) Informer [24] の注意機構の性能への寄与を調べるため、transformer 層を通常の transformer へと置き換えた。

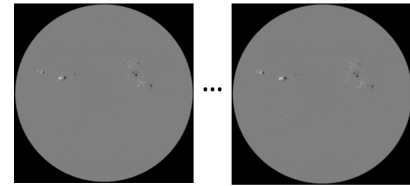
表 3 に ablation studies の定量的結果を示す。条件 (i) では、主要尺度である BSS および GMGS について、 $BSS_{\geq M}$ が 0.176 ポイント、GMGS が 0.06 ポイント減少した。したがって、cRT による再学習機構が十分に性能に寄与しているといえる。また条件 (ii), (iii), (iv) のどの条件も、3 つの尺度において同時に専門家を越える予測性能を達成



(a) 正しく予測した X クラスの太陽フレア



(b) 正しく予測した M クラスの太陽フレア



(c) 予測に失敗した例

図 3 定性的結果

表 3 Ablation studies の定量的結果

Method	$BSS_{\geq M} \uparrow$	GMGS $\uparrow$	$TSS_{\geq M} \uparrow$
(i) w/o cRT	0.046 ± 0.96	0.440 ± 0.21	0.574 ± 0.13
(ii-a) $(N_V, N_F) = (2, 2)$	0.26 ± 0.67	0.462 ± 0.19	0.500 ± 0.12
(ii-b) $(N_V, N_F) = (3, 3)$	0.24 ± 0.72	0.466 ± 0.20	0.511 ± 0.16
(iii) ConvNeXt $\rightarrow$ CNN	0.27 ± 0.68	0.473 ± 0.22	0.485 ± 0.17
(iv) Informer $\rightarrow$ Vanilla transformer	0.30 ± 0.64	0.544 ± 0.19	0.492 ± 0.12
提案手法	0.22 ± 0.68	0.500 ± 0.19	0.530 ± 0.11

した条件は存在しない。したがって、transformer 層の層数は $(N_V, N_F) = (1, 1)$ が最良であり、提案手法における transformer 層の構造、および ConvNeXt [9] が性能に寄与していることが確認できた。

5.5 エラー分析

最後に、エラー分析を行う。表 4 にて、2017 年のテスト集合における $GMGS_{\text{Influence}}$ [25] を示す。ただし $GMGS_{\text{Influence}}$ は以下の式で定義される。

$$GMGS_{\text{Influence}} = \frac{c_{ij}(s_{ii} - s_{ij})}{N} \quad (18)$$

ここで、 c_{ij} および s_{ij} はそれぞれ混同行列および GMGS のスコア行列 [2] の (i, j) 成分を示す。

表 4 より、ボトルネックは M クラスのフレアを誤って C クラスと予測する点であるといえる。したがって、このボトルネックを改善することがさらなる性能の向上につながると考えられる。

6 おわりに

本論文では、24 時間以内に発生する最大の太陽フレアクラスを予測するモデルとして、Flareformer を提案した。提案手法による貢献は以下である。

- 太陽フレア予測を扱う提案手法において、時系列予測に適した構造である Informer [25] の注意機構構造を導入した。
- 太陽フレア予測に寄与する特徴量を抽出するため、ConvNeXt [9] を導入した。
- サンプル数を伴う訓練時において、サンプル数の少ない X クラスへの過学習を防ぐため、Classifier Re-training (cRT) [4] に基づく分類器の再学習機構を導入した。
- BSS, GMGS, TSS において、提案手法は専門家を超える予測性能を達成した。

表4 2017 年テスト集合における GMGS_{Influence}

観測クラス	予測クラス	GMGS _{Influence}
M	C	0.0150
C	O	0.0113
O	C	0.0109
M	O	0.0090
O	M	0.0067
X	O	0.0054
C	X	0.0038
O	X	0.0023
M	X	0.0019
C	M	0.0018

謝辞

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

参考文献

- [1] George Box, Gwilym Jenkins, Gregory Reinsel et al., *Time Series Analysis: Forecasting and Control*, John Wiley & Sons (2015)
- [2] Lev Gandin et al., “Equitable Skill Scores for Categorical Forecasts”, *Monthly Weather Review*, Vol. 120, No. 2, pp. 361–370 (1992)
- [3] Rob Hyndman, Anne Koehler et al., *Forecasting with Exponential Smoothing: The State Space Approach*, Springer Science & Business Media (2008)
- [4] Bingyi Kang, Saining Xie, Marcus Rohrbach, Zhicheng Yan, Albert Gordo, Jiashi Feng et al., “Decoupling representation and classifier for long-tailed recognition”, in *ICLR* (2020)
- [5] Yuki 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)
- [6] Kanya Kusano, Tomoya Iju, Yumi Bamba et al., “A Physics-based Method that can Predict Imminent Large Solar Flares”, *Science*, Vol. 369, No. 6503, pp. 587–591 (2020)
- [7] James Lemen, David Akin et al., “The Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO)”, in *The Solar Dynamics Observatory*, pp. 17–40, Springer (2011)
- [8] Shiyang Li, Xiaoyong Jin, Yao Xuan et al., “Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting”, in *NeurIPS*, Vol. 32, pp. 5243–5253 (2019)
- [9] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell et al., “A convnet for the 2020s”, in *CVPR*, pp. 11976–11986 (2022)
- [10] Sophie Murray, Suzy Bingham, Michael Sharpe et al., “Flare forecasting at the met office space weather operations centre”, *Space Weather*, Vol. 15, No. 4, pp. 577–588 (2017)
- [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, Yuki Kubo, Mitsue Den 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 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, Jan asthaus, Lorenzo Stella, Yuyang Wang et al., “Deep State Space Models for Time Series Forecasting”, in *NeurIPS*, Vol. 31, pp. 7785–7794 (2018)
- [18] David Salinas, Valentin Flunkert 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 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 et al., “Informer: Beyond efficient transformer for long sequence time-series forecasting”, *AAAI*, Vol. 35, No. 12, pp. 11106–11115 (2021)
- [25] 兼田 寛大, 西塚 直人, 飯田 紘 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”, in *JSAI2021* (2021)