The Astrophysical Journal Letters, 890:L5 (7pp), 2020 February 10 https://doi.org/10.3847/2041-8213/ab701b

Forecast of Major Solar X-Ray Flare Flux Profiles Using Novel Deep Learning Models

Kangwoo Yi, Yong-Jae Moon, Gyungin Shin, and Daye Lim
School of Space Research, Kyung Hee university, 1732, Deogyeongdae-ro, Giheung-gu, Yongin-si Gyunggi-do, 17104, Republic of Korea; moonyj@khu.ac.kr
Received 2019 December 16; revised 2020 January 22; accepted 2020 January 24; published 2020 February 7

Abstract

In this Letter, we present the application of a couple of novel deep learning models to the forecast of major solar X-ray flare flux profiles. These models are based on a sequence-to-sequence framework using long short-term memory cell and an attention mechanism. For this, we use *Geostationary Operational Environmental Satellite* 10 X-ray flux data from 1998 August to 2006 April. Seven hundred sixty events are used for training and 85 for testing. The models forecast 30 minutes of X-ray flux profiles during the rise phase of the solar flare with a minute time cadence. We evaluate the models using the 10-fold cross-validation and rms error (RMSE) based on flux profiles and RMSE based on its peak flux. For comparison we consider two simple deep learning models and four conventional regression models. Major results of this study are as follows. First, we successfully apply our deep learning models to the forecast of solar flare X-ray flux profiles, without any preprocessing to extract features from data. Second, our proposed models outperform the other models. Third, our models achieve better performance for forecasting X-ray flux profiles with low-peak fluxes than those with high-peak fluxes. Fourth, our models successfully predict flare duration with high correlations for both all cases and cases at peak times. Our study indicates that our deep learning models can be useful for forecasting time-series data in astronomy and space weather, even for impulsive events such as major flares.

Unified Astronomy Thesaurus concepts: The Sun (1693); Solar x-ray flares (1816); Time series analysis (1916)

1. Introduction

Solar flare is one of the most energetic activities of the Sun, releasing a huge amount of energy in a broad spectrum of emission and accelerating particles from the Sun to the interplanetary space. An intense X-ray flux from large flare events affects spacecraft anomalies, the safety of astronauts, and radio fade-outs (Huang et al. 2018). These hazards can result in enormous economic losses (Siscoe 2000) but it is very difficult to pre-estimate the characteristics of solar flares such as peak flux or duration because solar radiation travels at the speed of light. National Oceanic and Atmospheric Administration (NOAA) classifies solar flares into five classes (A, B, C, M, and X) according to its peak soft X-ray flux (0.1–0.8 nm), and alerts radiation hazards when a major class flare (M and X) occurs.

There are several different types of flare forecasts. There have been many studies to predict solar flare occurrence probability within 24 hr: by statistical method (McIntosh 1990; Gallagher et al. 2002; Lee et al. 2012; Lim et al. 2019) and by machine learning methods (Colak & Qahwaji 2009; Song et al. 2009). Many other studies forecast solar flare occurrence within 24 hr with "Yes" or "No": by statistical methods (Bloomfield et al. 2012), by machine learning methods (Al-Ghraibah et al. 2015; Bobra & Couvidat 2015; Nishizuka et al. 2017; Park et al. 2017), and by deep learning (Huang et al. 2018; Nishizuka et al. 2018; Park et al. 2018; Liu et al. 2019). Several studies predict a maximum class of flare within 24 hr: by machine learning methods (Lee et al. 2007; Liu et al. 2017), and by deep learning (Hada-Muranushi et al. 2016). These studies are based on either McIntosh class or extracted magnetic properties of active regions. To the best of our knowledge, there is no study that forecasts the X-ray flux profiles of solar flares. Forecasting flare flux profiles, including the information of flare peak flux and duration, could help various space missions. For example, it would be helpful to assume the duration of ionospheric electron enhancement which could affect spacecraft operations and HF communication.

In this Letter, for the first time we present a couple of solar flare forecast models that predict the X-ray flux profiles of major solar flares in real-time with one minute cadence. These models are based on novel deep learning methods such as the sequence-to-sequence (seq2seq; Sutskever et al. 2014) framework using the Long Short-Term Memory (LSTM; Hochreiter & Schmidhuber 1997) and attention mechanism (Bahdanau et al. 2014), which is similar to Google Neural Machine Translation Network (Wu et al. 2016). We apply our models to the *Geostationary Operational Environmental Satellite* (*GOES*) X-ray flux profiles of 845 major flares. We compare our results with conventional regression models and different deep learning models. In contrast to the previous research based on active region analysis, we use historical solar flare soft X-ray flux data without any extracting features from data.

This paper is organized as follows. The data are described in Section 2. Our models are explained in Section 3.1. Results and analysis are given in Section 4. A brief conclusion and discussion are presented in Section 5.

2. Data

GOES observes solar X-ray fluxes in the two broadband channels of 1–8 Å and 0.5–4 Å. NOAA identifies solar flares automatically using one-minute-averaged *GOES* 1–8 Å X-ray fluxes. A flare is defined using the following algorithm (Ryan et al. 2016; NOAA GOES X-ray flux[1]): (1) the begin time of an X-ray event is defined as the first minute, in a sequence of 4 minutes, of steep monotonic increase in 0.1–0.8 nm flux, (2) the flux at the end of the fourth minute is at least 40% greater than the flux in the first minute, (3) the X-ray event peak time is

[1] https://www.swpc.noaa.gov/products/goes-x-ray-flux

taken as the minute of the peak X-ray flux, and (4) the end time is the time when the flux level decays to a point halfway between the maximum flux and the background level. We apply the algorithm to *GOES* 10 one-minute X-ray flux data, from 1998 August to 2006 May, to recognize major solar flares. In order to select reliable solar flare events, we remove multiple-peak events and pre-flare phase events by simple automatic process and visual inspection. As a result, we select 845 major solar flare events.

Soft X-ray fluxes from 26 minutes before flare start time to 3 minutes after it are used for input and those of following 30 minutes are used for target. Our data set is constructed using the sliding window method. The window slides from 3 minutes after flare start time to its peak time. We randomly split the data set into training and test sets since flare X-ray profiles can be assumed to be mostly independent one another in that flares have a Poisson distribution in time (Rosner & Vaiana 1978; Moon et al. 2001; ?, ?).

The data set is separated by each flare event to avoid that the data from the same flare event are used for training and test both. It is noted that when we divide the data set chronologically, the number of major flares in the test data set is too small. In order to reduce the effect of test data selection and random initial weights, we consider 10-fold cross-validation. As a result, 6227–6381 data (90% of events; 69 X-class and 691 M-class) are used for training and 654–808 data (10% of events; 8 X-class and 77 M-class) for test. The difference in data number is due to different lengths of flare events in each cross-validation set.

3. Model

3.1. Methods

LSTM is one of the advanced recurrent neural networks (Hopfield 1982) that is good to learn time dependence in sequence. LSTM have been widely employed in various fields such as machine translation (Sutskever et al. 2014; Wu et al. 2016), speech recognition (Graves & Schmidhuber 2005), and time-series prediction (Duan et al. 2018; Tan et al. 2018). LSTM layer has a cell state candidate and three gates, to transfer hidden state from prior time step to next time step. The time step means a sequential order of elements in time-series data. Details of the LSTM calculation are described below.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + B_f), \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + B_i), \quad (2)$$

$$\widetilde{c_t} = \tanh(W_c \cdot [h_{t-1}, x_t] + B_c), \quad (3)$$

$$c_t = f_t \odot C_{t-1} + i_t \odot \widetilde{c_t}, \quad (4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + B_o), \quad (5)$$

$$h_t = o_t \odot \tanh(c_t), \quad (6)$$

where f_t represents the forget gate, i_t represents the input gate, $\widetilde{c_t}$ represents the cell state candidate, c_t represents the cell state, o_t represents the output gate, and h_t represents the hidden state at time step t; X is input data; W terms represent weight metrics and B terms are bias, which are trainable; σ is a logistic sigmoid function and tanh is a hyperbolic tangent function. Here $\cdot$ denotes dot product and $\odot$ denotes element-wise multiplication. The forget gate, input gate, and output gate calculate weights that are applied to past information, new information, and output corresponding to forecasting result, respectively. The cell state candidate is modified information considering how much input information should be reflected to the cell state. The hidden state is a result of the LSTM layer.

Seq2seq framework relies two LSTM models: one is an encoder for extracting information from input data and the other is a decoder for making a forecast using the information from the encoder. The decoder repeats a process to generate a prediction and then to use the predicted result for the next prediction. This mechanism has been used for generating time-series data (Wu et al. 2016; Naul et al. 2018).

In addition, we employ an attention mechanism that allows the model to learn and focus on important time steps of input data. The attention computes a context vector c_i corresponding to the weighted sum of source state h,

$$c_i = \sum_{j=1}^{T_x} W_{ij} h_j, \quad (7)$$

where T_x represents length of input and W represents weight. Here i denotes time step of output and j denotes time step of input. The weight W is computed by

$$W_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})}, \quad (8)$$

$$e_{ij} = a(s_{i-1}, h_j), \quad (9)$$

where s is a hidden state. Here a denotes an alignment model that calculates a score reflecting how well output at the position i and input at the position j match each other. We adopt the additive attention model (Bahdanau et al. 2014)

$$a(s_{i-1}, h_j) = v^T \tanh(W_s s_{i-1} + W_h h_j), \quad (10)$$

where v and W are learnable parameters.

3.2. Architecture

Figure 1 shows an overall architecture of our first proposed model. This model takes two LSTM layers and one fully connected layer for the encoder and decoder. We adopt a fully connected layer with the ReLU (Nair & Hinton 2010) activation function for the encoder and a fully connected layer without an activation function for the decoder. The second model is obtained by removing the attention layer from Figure 1. The mean-squared loss function and Adam optimizer (Kingma & Ba 2014) with learning rate 0.001 are used for our models.

4. Results

To evaluate models, we consider three types of rms errors (RMSEs), one for all forecasting results, another for peak flux, and the other for different forecast start time:

$$RMSE_{\text{all}} = \sqrt{\frac{\sum_{e=1}^{E}\sum_{i=3}^{p}\sum_{j=1}^{30}(\log(y_{ei}^{j}) - \log(f_{ei}^{j})^2}{N_{\text{all}}}}, \quad (11)$$

$$RMSE_{\text{peakflux}} = \sqrt{\frac{\sum_{e=1}^{E}\sum_{i=3}^{p}(\max(\log(y_{ei}^{j})_{j=1}^{30} - \max(\log(f_{ei}^{j})_{j=1}^{30})^2}{N_{\text{peakflux}}}}, \quad (12)$$

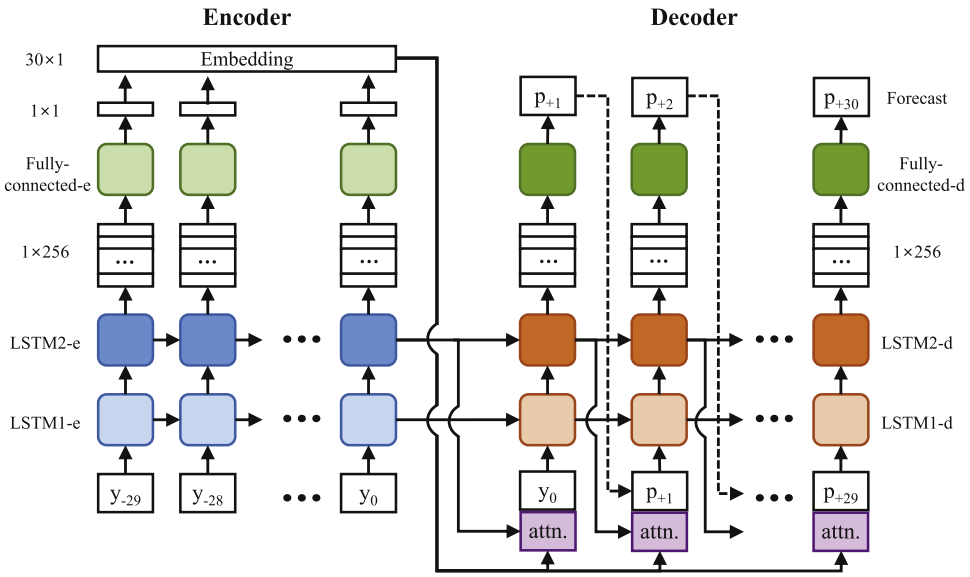


Figure 1. Architecture of our first proposed model. Two LSTM layers and one fully connected layer for the encoder and decoder. The encoder takes 30 minutes of X-ray flux (y) and produces the embedded information constructed by passing the output of LSTM layers into a fully connected layer. The decoder repeats a similar process with the encoder 30 times. In every step, the decoder takes a result of the attention layer and new information to predict X-ray flux (p) of the next 1 minute which is used for next step prediction corresponding to new information. Results of LSTM layer in decoder are sent to a fully connected layer with a linear activation function. The second proposed model has the same structure without attention.

Table 1
Results of the Models

Model	$\mathrm{RMSE_{all}}$	$\mathrm{RMSE_{peakflux}}$	RMSE_t				Hyperparameters
			$t = 3$	$t = 4$	$t = 5$	$t = 6$	
ARIMA	0.72 ± 0.11	0.63 ± 0.19	0.74 ± 0.09	1.20 ± 0.35	0.86 ± 0.14	0.74 ± 0.14	p: 1–5, d: 1, q: 1–5, Input data length: 1800
KNNR	0.39 ± 0.02	0.31 ± 0.03	0.45 ± 0.02	0.44 ± 0.02	0.44 ± 0.02	0.43 ± 0.02	K: 58, Metric: Euclidean
RFR	0.38 ± 0.02	0.30 ± 0.03	0.45 ± 0.02	0.44 ± 0.02	0.43 ± 0.02	0.42 ± 0.02	Trees: 100, M:1.0 , Min sample split: 0.025
SVMR	0.37 ± 0.02	0.29 ± 0.03	0.46 ± 0.03	0.45 ± 0.03	0.42 ± 0.02	0.39 ± 0.03	Kernel: RBF, C: 1, epsilon: 0.005
MLP	0.45 ± 0.02	0.45 ± 0.05	0.52 ± 0.04	0.50 ± 0.04	0.48 ± 0.03	0.46 ± 0.02	Fully connected layers: 18 (30–280, 280–280, ⋯, 280–30)
Simple-LSTM	0.50 ± 0.05	0.40 ± 0.05	0.53 ± 0.02	0.54 ± 0.02	0.53 ± 0.02	0.52 ± 0.03	Hidden units: 48, Fully connected layers: 2 (96–96, 96–1)
Seq2seq	0.33 ± 0.02	0.26 ± 0.03	0.42 ± 0.02	0.41 ± 0.02	0.38 ± 0.02	0.36 ± 0.02	Hidden units: 192, Embedding size: 8, Fully connected layers: 1(192–1)
Seq2seq+attention	0.32 ± 0.02	0.26 ± 0.02	0.41 ± 0.02	0.40 ± 0.02	0.37 ± 0.02	0.35 ± 0.02	Hidden units: 256, Embedding size: 1, Fully connected layers: 1(256–1)

Note. RMSEs are expressed with standard deviations of 10-fold cross-validation trials. Fully connected layer is described as input size–output size.

$$RMSE_t = \sqrt{\frac{\sum_{e=1}^{E}\sum_{j=1}^{30}(\log(y_{ei}^{j})|_{i=t} - \log(f_{ei}^{j})|_{i=t})^2}{N_t}}, \quad (13)$$

where y represents an observed flux and f represents a forecasted flux. N_{all}, N_{peakflux}, and N_t are the numbers of all cases, at peak times and at times t after flare start, respectively. Here E, i, j, and p are the number of flares, forecast start time, each forecasting time, and peak time, respectively. This process was executed 10 times for each cross-validation test data set and the average of the results are shown.

For comparison with our models, we make two simple deep learning models and four conventional regression models. The first deep learning model is based on Multi-Layer-Perceptron (MLP), constructed of several fully connected layers. This model analyzes data as independent features, not time-series features. The second deep learning model is based on LSTM, constructed of two LSTM layers and two fully connected layer. This model can analyze data as time-series features. These models are trained using the same loss function and optimizer with our models. Four conventional models are based on the Auto-Regressive Integrated Moving Average (ARIMA; Luceño & Peña 2008), the K-Nearest Neighbor Regression (KNNR; Altman 1992), the Support Vector Machine Regression (SVMR; Cortes & Vapnik 1995), and the Random Forest Regression (RFR; Breiman 2001). For the ARIMA model, we

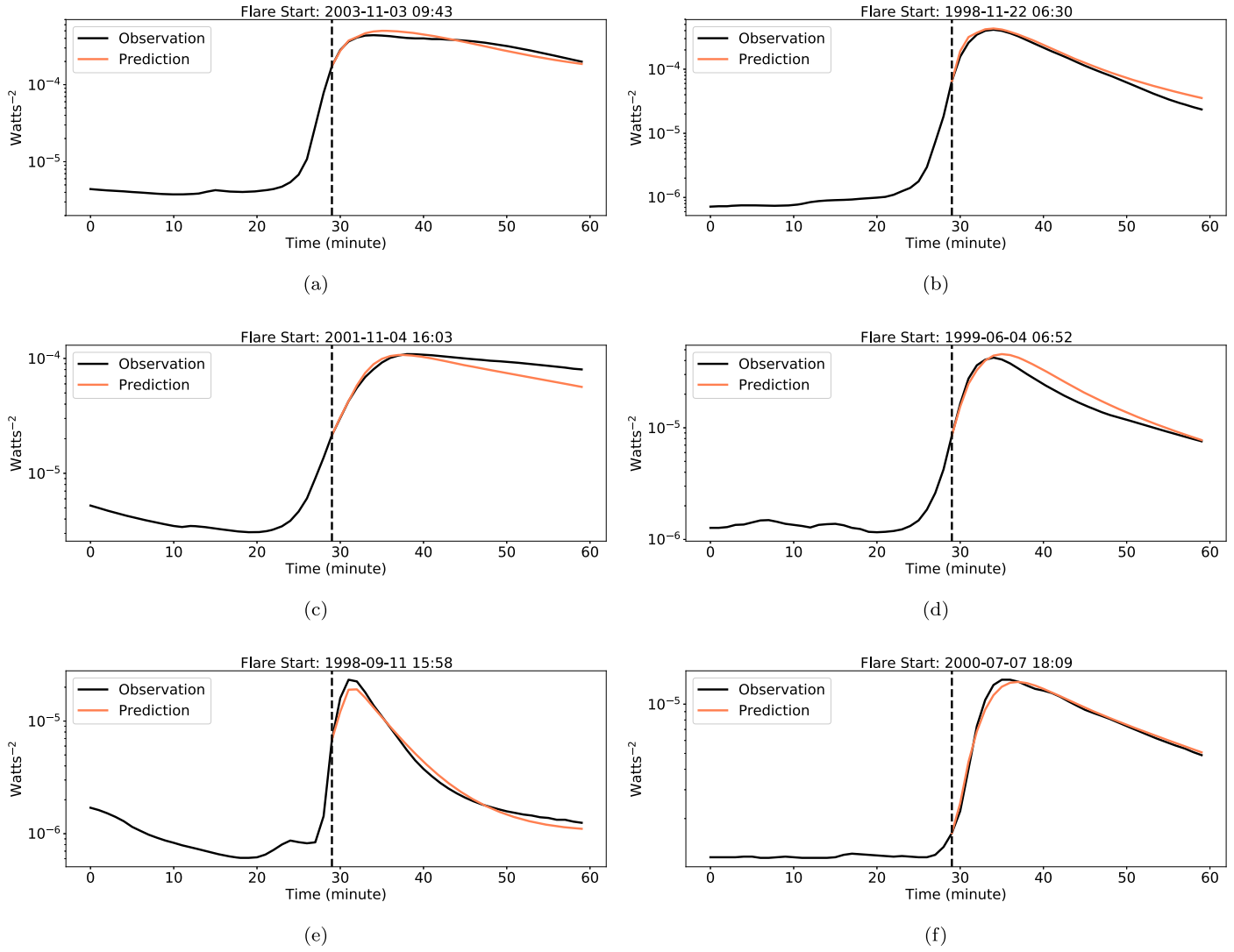


Figure 2. Results of the seq2seq+attention model. Panels (a), (b), (c), (d), (e), and (f) are the X5, X4, X1, M4, M2, and M1 flare forecasts at 7, 7, 7, 6, 3, and 3 minutes after flare start time, respectively. 0–29 minutes are input data and 30–59 minutes are output data together with observations. The black vertical dashed lines denote forecast start times.

use the AUTO-ARIMA process (Hyndman & Khandakar 2008) for all input data to find the best results of the ARIMA model.

For comparing the models fairly, we test many hyperparameters for conventional regression models and deep learning models (48, 96, 128, 256, and 512 hidden units and nodes; 1, 8, 16 embedding and no embedding; and 2–18 fully connected layers with a residual connection; He et al. 2015), and carry out the early stopping (Morgan & Bourlard 1990) process during the deep learning model training to evade the overfitting problem and find the best result of each deep learning model.

Table 1 shows the result of the models. Considering the standard deviation of 10-fold cross-validation, our proposed models (seq2seq, seq2seq+attention) outperform the others in all metrics. The differences between our proposed models are not significant. The seq2seq model could be better than the seq2seq+attention model depending on the data set. For more discussion, we select the seq2seq+attention model for forecast analysis because it produces the best results.

For the seq2seq+attention model, $\mathrm{RMSE_{all}}$ and $\mathrm{RMSE_{peakflux}}$ are 0.32 and 0.26, which are much better than those from the four regression models and two simple deep learning models. The RMSE_t, which depends on forecast start time (3–6 minutes), is also better for proposed models than those for other models. Our results demonstrate that adopting both seq2seq and attention are effective for forecasting the solar X-ray profiles. In all eight models, the results of the later forecasts (RMSE_t with higher t) are better than the ones of the earlier forecasts (RMSE_t with lower t) because of rapid change of the X-ray flux profiles.

In addition, we divide forecasts of the seq2seq+attention model into three groups according to flare class (M1–M4.9, M5–M9.9, and X1-). As a result, $\mathrm{RMSE_{all}}$ and $\mathrm{RMSE_{peakflux}}$ are 0.27 and 0.15 for the M1–M4.9 flare, 0.36 and 0.31 for the M5–M9.9 flare, and 0.51 and 0.53 for the X1- flare. This analysis indicates that strong solar flares are harder to predict than weak solar flares.

Figure 2 is the forecast results of the seq2seq+attention model. The model well forecasts flare peak flux, peak time, and overall profiles.

Figure 3 is the forecasts of the seq2seq+attention model from 5 minutes after flare start to peak time for the X5-class flare shown in Figure 2(a). As shown in Figure 3(a), its peak flux is noticeably underestimated. However, when the *GOES* observation is at the M class, our model forecasts that the flare

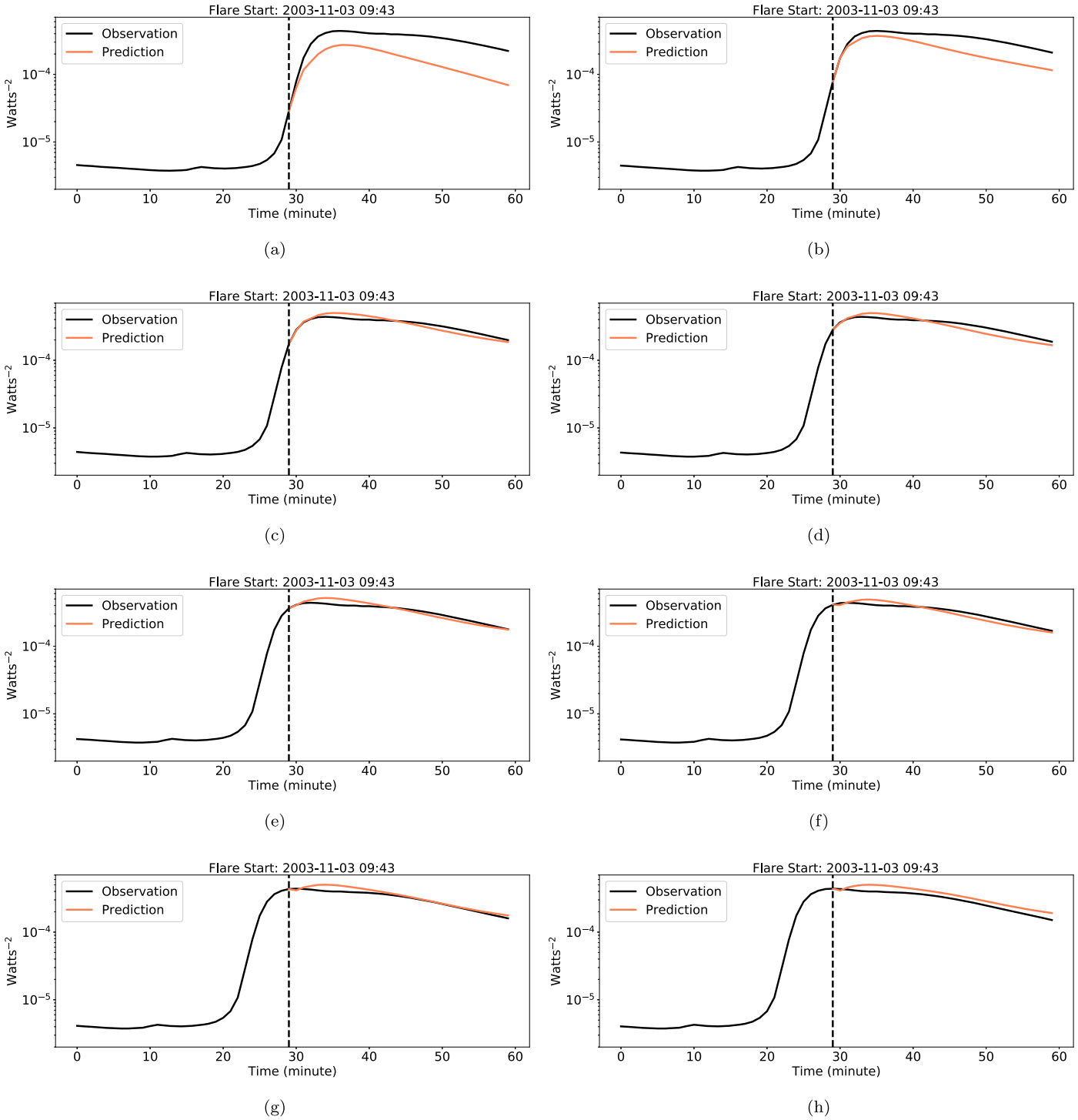
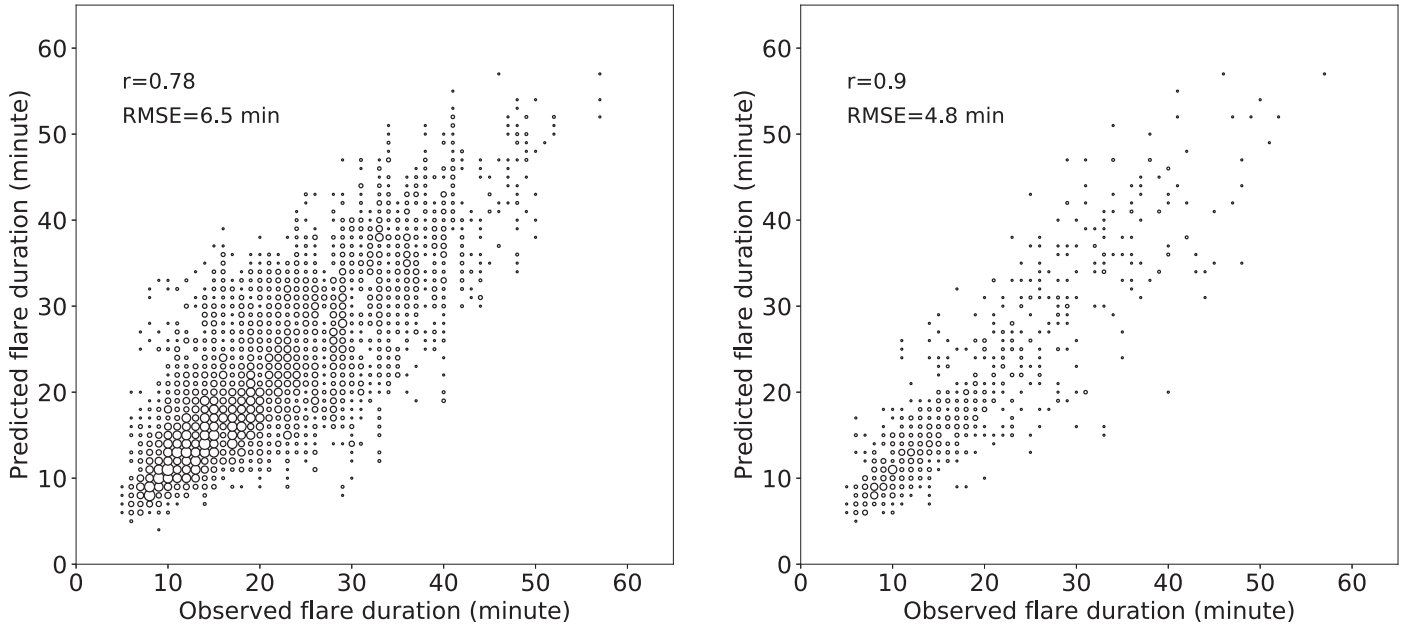


Figure 3. Prediction of the seq2seq+attention model for an X5-class flare on 2003 November 3. Panels (a)–(h) show the forecasts at 5–12 minutes after the flare start time, in chronological order. 0–29 minutes are the input and 30–59 minutes are the output together with observations. The black vertical dashed lines denote forecast start times.

will reach an X class. In Figure 3(b), our model predicts a peak flux similar to observation. In subsequent forecasts (Figures 3(c)–(h)), the model well predicts its peak flux and X-ray flux profiles in the decay phase.

Figure 4 shows a relationship between predictions and observations of flare duration. To evaluate the flare duration prediction of the seq2seq+attention model, we calculate the correlation coefficient and RMSE in the case that both flare end times for observation and prediction are within 30 minutes. Because of the forecasting period, the data points that have no observed and/or no predicted flare end time within 30 minutes cannot be displayed on the scatter plot. The correlation coefficient and RMSE are 0.78 and 6.5 minutes for all forecasts, while they are 0.9 and 4.8 minutes for the peak time forecast. These results show that our model well forecasts flare duration for both cases. It is also noted that the prediction of flare duration at the peak time is very excellent in view of correlation and RMSE.



(a) (b)

Figure 4. Predicted flare duration vs. observed flare duration for (a) all forecasts and (b) forecasts at peak times by the seq2seq+attention model. Size of the circles denotes the number of predictions.

5. Conclusion and Discussion

In this Letter, we have presented new forecast models to predict X-ray flux profiles of solar major flares. These novel deep learning models are based on seq2seq with/without attention. We take *GOES* 10 X-ray fluxes of 845 major flare events from 1998 August to 2006 May. We randomly separate our data sets into 90% for training and 10% for test. We train and evaluate our models using 10-fold cross-validation, and compare our results with other deep learning models and conventional regression models. The main results from this study are summarized as follows. First, we successfully apply our models to the forecast of solar flare X-ray flux profiles, without any preprocessing to extract features from the data. Second, in view of RMSE, our models outperform other deep learning models and conventional regression models with 0.32 and 0.33 for $\mathrm{RMSE_{all}}$ and 0.26 for $\mathrm{RMSE_{peakflux}}$. Third, $\mathrm{RMSE_{all}}$ and $\mathrm{RMSE_{peakflux}}$ of the seq2seq+attention model are 0.27 and 0.15 for the M1–M4.9 flare, 0.36 and 0.31 for the M5–M9.9 flare, and 0.51 and 0.53 for the X1- flare, respectively. Fourth, in flare duration prediction, the correlation coefficient and RMSE values from the seq2seq+attention model are 0.78 and 6.5 minutes for all forecasts, while they are 0.9 and 4.8 minutes for the peak time forecast.

We build the data set based on the assumption that the flares have a Poisson distribution in time (Rosner & Vaiana 1978; Moon et al. 2001; Wheatland 2000; Wheatland & Litvinenko 2002). On the other hand, there are different opinions such as a power-law distribution (Boffetta et al. 1999) and a Lévy distribution (Lepret et al. 2001). Data set construction could be affected by flare distribution in time.

Our model produces better predictions for weaker flares than stronger flares. This may be mainly caused by two reasons. The first reason is the class imbalance problem: the number of stronger flares is much less than that of weaker flares. In this case, the models are trained to focus on better predicting weaker flares than stronger flares. The second reason is that stronger flares have more dynamic X-ray profiles than weaker flares. This is already noted in Section 4 that $\mathrm{RMSE_{peakflux}}$ rapidly increases for stronger flares.

Recently, several deep learning methods have been applied to time-series data in astronomy and space weather. Muthukrishna et al. (2019) classified the transient class using a gated recurrent unit (Chung et al. 2014). By using the seq2seq model, Naul et al. (2018) reconstructed the light curve of variable stars and classified the class of variable stars better than a conventional algorithm, and Shen et al. (2017) generated denoised gravitational-wave signals.

In this work, we have demonstrated that solar X-ray profiles are well predicted by novel deep learning models that are based on seq2seq with LSTM. It is impressive that our models can well predict rapidly increasing and then decreasing patterns such as solar flares. We expect that our method can be applied to many time-series data in astronomy and space weather, even in other scientific fields. There seems to be several candidates for application of time-series data: variable stars (Chung et al. 2014; Naul et al. 2018), gravitational waves (Shen et al. 2017), and supernova evolution (Suwa 2017). On the other hand, our models can be useful for ionosphere behavior prediction such as D-region enhancement (McRae & Thomson 2004) or an increase in total electron content (Liu et al. 2004) since it can predict X-ray flux profiles in advance.

We appreciate the referee's constructive comments. This work was supported by the BK21 plus program through the National Research Foundation (NRF) funded by the Ministry of Education of Korea, the Basic Science Research Program through the NRF funded by the Ministry of Education (NRF-2016R1A2B4013131, NRF-2019R1A2C1002634), the Korea Astronomy and Space Science Institute (KASI) under the R&D program "Study on the Determination of Coronal Physical

Quantities using Solar Multi-wavelength Images" (project No. 2019-1-850-02) supervised by the Ministry of Science and ICT, and Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (2018-0-01422, Study on analysis and prediction technique of solar flares). The *GOES* X-ray data used here may be accessed through https://satdat.ngdc.noaa.gov/sem/goes/data/. We thank contributors to Pytorch, Numpy, Scikit-learn, Pmdarima, and Matplotlib open-source packages.

ORCID iDs

Kangwoo Yi https://orcid.org/0000-0003-4342-9483
Yong-Jae Moon https://orcid.org/0000-0001-6216-6944
Daye Lim https://orcid.org/0000-0001-9914-9080

References

Al-Ghraibah, A., Boucheron, L. E., & McAteer, R. T. J. 2015, A&A, 579, A64
Altman, N. S. 1992, The American Statistician, 46, 175
Bahdanau, D., Cho, K., & Bengio, Y. 2014, arXiv:1409.0473
Bloomfield, D. S., Higgins, P. A., McAteer, R. T. J., & Gallagher, P. T. 2012, ApJL, 747, L41
Bobra, M. G., & Couvidat, S. 2015, ApJ, 798, 135
Boffetta, G., Carbone, V., Giuliani, P., Veltri, P., & Vulpiani, A. 1999, PhRvL, 83, 4662
Breiman, L. 2001, Machine Learning, 45, 5
Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. 2014, arXiv:1412.3555
Colak, T., & Qahwaji, R. 2009, SpWea, 7, S06001
Cortes, C., & Vapnik, V. 1995, Machine Learning, 20, 273
Duan, Z., Yang, Y., Zhang, K., Ni, Y., & Bajgain, S. 2018, IEEE Access, 6, 31820
Gallagher, P. T., Moon, Y. J., & Wang, H. 2002, SoPh, 209, 171
Graves, A., & Schmidhuber, J. 2005, NN, 18, 602
Hada-Muranushi, Y., Muranushi, T., Asai, A., et al. 2016, arXiv:1606.01587
He, K., Zhang, X., Ren, S., & Sun, J. 2015, arXiv:1512.03385
Hochreiter, S., & Schmidhuber, J. 1997, Neural Computation, 9, 1735
Hopfield, J. J. 1982, PNAS, 79, 2554
Huang, X., Wang, H., Xu, L., et al. 2018, ApJ, 856, 7
Hyndman, R., & Khandakar, Y. 2008, Journal of Statistical Software, 27, 1
Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
Lee, J. Y., Moon, Y. J., Kim, K. S., Park, Y. D., & Fletcher, A. r. 2007, JKAS, 40, 99
Lee, K., Moon, Y. J., Lee, J.-Y., Lee, K.-S., & Na, H. 2012, SoPh, 281, 639
Lepret, F., Carbone, V., & Veltri, P. 2001, ApJL, 555, L133
Lim, D., Moon, Y.-J., Park, J., et al. 2019, JKAS, 52, 133
Liu, C., Deng, N., Wang, J. T. L., & Wang, H. 2017, ApJ, 843, 104
Liu, H., Liu, C., Wang, J. T. L., & Wang, H. 2019, ApJ, 877, 121
Liu, J. Y., Lin, C. H., Tsai, H. F., & Liou, Y. A. 2004, JGRA, 109, A01307
Luceño, A., & Peña, D. 2008, in Encyclopedia of Statistics in Quality and Reliability, ed. F. Ruggieri, R. S. Kennett, & F. W. Faltin (New York: Wiley),
McIntosh, P. S. 1990, SoPh, 125, 251
McRae, W. M., & Thomson, N. R. 2004, JASTP, 66, 77
Moon, Y. J., Choe, G. S., Yun, H. S., & Park, Y. D. 2001, JGR, 106, 29951
Morgan, N., & Bourlard, H. 1990, in Advances in Neural Information Processing Systems, ed. D. S. Touretzky, Vol. 2 (Burlington, MA: Morgan-Kaufmann), 630
Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., & Hložek, R. 2019, PASP, 131, 118002
Nair, V., & Hinton, G. E. 2010, in Proc. 27th Int. Conf. Machine Learning, ICML'10 (Madison, WI: Omnipress), 807, http://dl.acm.org/citation.cfm?id=3104322.3104425
Naul, B., Bloom, J. S., Pérez, F., & van der Walt, S. 2018, NatAs, 2, 151
Nishizuka, N., Sugiura, K., Kubo, Y., et al. 2017, ApJ, 835, 156
Nishizuka, N., Sugiura, K., Kubo, Y., Den, M., & Ishii, M. 2018, ApJ, 858, 113
Park, E., Moon, Y.-J., Shin, S., et al. 2018, ApJ, 869, 91
Park, J., Moon, Y.-J., Choi, S., et al. 2017, SpWea, 15, 704
Rosner, R., & Vaiana, G. S. 1978, ApJ, 222, 1104
Ryan, D. F., Dominique, M., Seaton, D., Stegen, K., & White, A. 2016, A&A, 592, A133
Shen, H., George, D., Huerta, E. A., & Zhao, Z. 2017, arXiv:1711.09919
Siscoe, G. 2000, JASTP, 62, 1223
Song, H., Tan, C., Jing, J., et al. 2009, SoPh, 254, 101
Sutskever, I., Vinyals, O., & Le, Q. V. 2014, arXiv:1409.3215
Suwa, Y. 2017, MNRAS, 474, 2612
Tan, Y., Hu, Q., Wang, Z., & Zhong, Q. 2018, SpWea, 16, 406
Wheatland, M. S. 2000, ApJL, 536, L109
Wheatland, M. S., & Litvinenko, Y. E. 2002, SoPh, 211, 255
Wu, Y., Schuster, M., Chen, Z., et al. 2016, arXiv:1609.08144