Predicting Solar Flares Using a Long Short-term Memory Network

Hao Liu, Chang Liu, Jason T. L. Wang and Haimin Wang

1. Introduction

Solar flares, the largest explosive events in our solar system, are intense bursts of radiation that occur in the Sun's atmosphere and release massive amounts of energy into space. They last from minutes to hours and are often seen as bright chromospheric ribbons and hot coronal loops on the Sun. Some flares are small and innocent while others can be tremendous and violent. Powerful flares and the often accompanied coronal mass ejections (CMEs) can cause severe influences on the near-Earth environment, resulting in potentially life-threatening consequences (Daglis et al. 2004). Therefore, substantial efforts are being invested on solar-flare research including forecast and mitigation plans.

In this notebook, we attempt to use SDO/HMI vector magnetic field data together with flaring history to predict solar flares that would occur in an AR within 24 hr of a given time point, with a deep-learning method, named long short-term memory (LSTM; Hochreiter & Schmidhuber 1997).

2. LSTM Workflow

2.1 Data Prepration & Loading

The data folder includes three sub-directories: LSTM_C_sample_run, LSTM_M_sample_run, and LSTM_M5_sample_run.

- The LSTM_C_sample_run includes a CSV training data file that is used to train the model and a CSV test data file that is used to predict the C category and higher flares.
- The LSTM_M_sample_run includes a CSV training data file that is used to train the model and a CSV test data file that is used to predict the M category and higher flares.
- The LSTM_M5_sample_run includes a CSV training data file that is used to train the model and a CSV test data file that is used to predict the M5 category and higher flares.

The files are loaded and used during the testing and training process.

2.2 C Flare Model Training and Testing

You may train the model with your own data or train the model with the default data (see Sections 2.2.1 and 2.2.2).

2.2.1 C Flare Model Training with Default Data

Here, we show how to train the model with default data. To train the model with your own data:

1. You should first upload your file to the data directory (in the left hand side file list).
2. Edit the path to the training file:
   'train_data_file':'data/LSTM_C_sample_run/normalized_training.csv'
   and replace the value 'data/LSTM_C_sample_run/normalized_training.csv' with your new file name.

In [1]:
```
!pip freeze
```

In [2]:
```
print('Loading the train_model function...')
from flarepredict_train import train_model
args = {'train_data_file':'data/LSTM_C_sample_run/normalized_training.csv',
        'flare': 'C',
       'modelid': 'custom_model_id'
      }
train_model(args)
```

```
Loading the train_model function...

d:\jupyternotebooks\master_project\py36\lib\site-packages\tensorflow\python\fra
mework\dtypes.py:523: FutureWarning: Passing (type, 1) or '1type' as a synonym
of type is deprecated; in a future version of numpy, it will be understood as
(type, (1,)) / '(1,)type'.
  _np_qint8 = np.dtype([("qint8", np.int8, 1)])
d:\jupyternotebooks\master_project\py36\lib\site-packages\tensorflow\python\fra
mework\dtypes.py:524: FutureWarning: Passing (type, 1) or '1type' as a synonym
of type is deprecated; in a future version of numpy, it will be understood as
(type, (1,)) / '(1,)type'.
  _np_quint8 = np.dtype([("quint8", np.uint8, 1)])
d:\jupyternotebooks\master_project\py36\lib\site-packages\tensorflow\python\fra
mework\dtypes.py:525: FutureWarning: Passing (type, 1) or '1type' as a synonym
of type is deprecated; in a future version of numpy, it will be understood as
(type, (1,)) / '(1,)type'.
  _np_qint16 = np.dtype([("qint16", np.int16, 1)])
d:\jupyternotebooks\master_project\py36\lib\site-packages\tensorflow\python\fra
mework\dtypes.py:526: FutureWarning: Passing (type, 1) or '1type' as a synonym
of type is deprecated; in a future version of numpy, it will be understood as
(type, (1,)) / '(1,)type'.
  _np_quint16 = np.dtype([("quint16", np.uint16, 1)])
d:\jupyternotebooks\master_project\py36\lib\site-packages\tensorflow\python\fra
mework\dtypes.py:527: FutureWarning: Passing (type, 1) or '1type' as a synonym
of type is deprecated; in a future version of numpy, it will be understood as
(type, (1,)) / '(1,)type'.
  _np_qint32 = np.dtype([("qint32", np.int32, 1)])
d:\jupyternotebooks\master_project\py36\lib\site-packages\tensorflow\python\fra
mework\dtypes.py:532: FutureWarning: Passing (type, 1) or '1type' as a synonym
of type is deprecated; in a future version of numpy, it will be understood as
(type, (1,)) / '(1,)type'.
  np_resource = np.dtype([("resource", np.ubyte, 1)])
Using TensorFlow backend.

turn off loggins is not supported
Starting training with a model with id: custom_model_id training data file: dat
a/LSTM_C_sample_run/normalized_training.csv
Loading data set...
(84577, 10, 14)
Done loading data...
Training is in progress, please wait until it is done...

Finished training the C flare model, you may use the flarepredict_test.py progr
am to make prediction.
```

2.2.2 Predicting with Your C Flare Model

To predict the testing data using the model you trained above, make sure the modelid value in the args variable in the following code is set exactly as the one used in the training, for example: 'custom_model_id'.

In [3]:
```
from flarepredict_test import test_model
args =  {'test_data_file': 'data/LSTM_C_sample_run/normalized_testing.csv',
         'flare': 'C',
         'modelid': 'custom_model_id'}
result_file_name = test_model(args)
```

```
turn off loggins is not supported
Starting testing with a model with id: custom_model_id testing data file: data/
LSTM_C_sample_run/normalized_testing.csv
Loading data set...
(185, 10, 14)
Done loading data...
Formatting and mapping the data...
Prediction is in progress, please wait until it is done...
Finished the prediction task..
```

2.2.3 Reading the Results

The prediction result can be shown by passing the result variable to the pandas function read_csv as shown in the following example. The result shows the label your model predicted for each case.

In [4]:
```
import pandas as pd
df = pd.read_csv(result_file_name)
df
```

Out[4]:

	Predicted Label	label	flare	timestamp	NOAA	HARP	TOTUSJH	Cdec	TOTUSJZ	C
0	Negative	Negative	B2.8	2017-03-25T01:22:36.70Z	12644	6972	-0.805765	0.000000	-0.509193	0.0
1	Negative	Negative	B2.8	2017-03-25T02:22:36.70Z	12644	6972	-0.769690	0.000000	-0.473642	0.0
2	Negative	Negative	B2.8	2017-03-25T04:22:36.80Z	12644	6972	-0.755696	0.000000	-0.441264	0.0
3	Negative	Negative	B2.8	2017-03-25T15:22:36.60Z	12644	6972	-0.713604	0.000000	-0.505940	0.0
4	Negative	Negative	B2.8	2017-03-25T16:22:36.60Z	12644	6972	-0.681474	0.000000	-0.348887	0.0
...	...	...	...	...	...	...	...	...	...	
180	Positive	Positive	M5.3	2017-04-02T07:22:37.90Z	12644	6972	1.735370	0.126305	1.325002	0.
181	Positive	Positive	M2.3	2017-04-02T08:22:37.90Z	12644	6972	1.810731	0.116207	1.202598	0.
182	Positive	Positive	M2.3	2017-04-02T09:22:37.90Z	12644	6972	1.799658	0.106915	1.588251	0.
183	Positive	Positive	M2.3	2017-04-02T10:22:37.90Z	12644	6972	1.819823	0.098367	1.659994	0.
184	Positive	Positive	M2.3	2017-04-02T11:22:37.90Z	12644	6972	1.968782	0.090502	1.791367	0.

185 rows × 20 columns

2.3 M Flare Model Training and Testing

You may train the model with your own data or train the model with the default data (see Sections 2.3.1 and 2.3.2).

2.3.1 M Flare Model Training with Default Data

Here, we show how to train the model with default data. To train the model with your own data:

1. You should first upload your file to the data directory (in the left hand side file list).
2. Edit the path to the training file:
   'train_data_file':'data/LSTM_M_sample_run/normalized_training.csv'

and replace the value 'data/LSTM_M_sample_run/normalized_training.csv' with your new file name.

In [5]:
```
print('Loading the train_model function...')
from flarepredict_train import train_model
args = {'train_data_file':'data/LSTM_M_sample_run/normalized_training.csv',
        'flare': 'M',
       'modelid': 'custom_model_id'
      }
train_model(args)
```

```
Loading the train_model function...
turn off loggins is not supported
Starting training with a model with id: custom_model_id training data file: dat
a/LSTM_M_sample_run/normalized_training.csv
Loading data set...
(84742, 10, 22)
Done loading data...
Training is in progress, please wait until it is done...

Finished training the M flare model, you may use the flarepredict_test.py progr
am to make prediction.
```

2.3.2 Predicting with Your M Flare Model

To predict the testing data using the model you trained above, make sure the modelid value in the args variable in the following code is set exactly as the one used in the training, for example: 'custom_model_id'.

In [6]:
```
from flarepredict_test import test_model
args =  {'test_data_file': 'data/LSTM_M_sample_run/normalized_testing.csv',
         'flare': 'M',
         'modelid': 'custom_model_id'}
result_file_name = test_model(args)
```

```
turn off loggins is not supported
Starting testing with a model with id: custom_model_id testing data file: data/
LSTM_M_sample_run/normalized_testing.csv
Loading data set...
(212, 10, 22)
Done loading data...
Formatting and mapping the data...
Prediction is in progress, please wait until it is done...
Finished the prediction task..
```

2.3.3 Reading the Results

The prediction result can be shown by passing the result variable to the pandas function read_csv as shown in the following example. The result shows the label your model predicted for each case.

In [7]:
```
import pandas as pd
df = pd.read_csv(result_file_name)
df
```

Out[7]:

	Predicted Label	Label	flare	timestamp	NOAA	HARP	TOTUSJH	TOTUSJZ	TOTPOT	T
0	Negative	Negative	C1.0	2015-04-04T05:34:17.60Z	12320	5415	-0.104048	-0.301538	-0.137916	-0.
1	Negative	Negative	C1.0	2015-04-04T06:34:17.60Z	12320	5415	-0.078954	-0.175091	-0.094344	-0.
2	Negative	Negative	C1.0	2015-04-04T07:34:17.60Z	12320	5415	-0.043782	-0.338201	-0.018729	-0.
3	Negative	Negative	C1.0	2015-04-04T08:34:17.60Z	12320	5415	0.018644	-0.306688	0.096848	-0.
4	Negative	Negative	C1.0	2015-04-04T09:34:17.60Z	12320	5415	0.055247	-0.372511	0.116056	-0.
...	...	...	...	...	...	...	...	...	...	
207	Positive	Negative	C2.9	2015-04-13T15:34:18.70Z	12320	5415	0.947701	-0.229454	0.790858	0.
208	Positive	Negative	C2.9	2015-04-13T16:34:18.70Z	12320	5415	0.945490	-0.256896	0.851387	0.
209	Positive	Negative	C2.9	2015-04-13T17:34:18.70Z	12320	5415	0.826441	-0.314707	0.766035	0.
210	Positive	Negative	C2.9	2015-04-13T18:34:18.70Z	12320	5415	0.944913	-0.175675	0.811451	0.
211	Positive	Negative	C2.9	2015-04-13T19:34:18.80Z	12320	5415	0.959147	-0.067101	0.905028	0.

212 rows × 28 columns

2.4 M5 Flare Model Training and Testing

You may train the model with your own data or train the model with the default data (see Sections 2.4.1 and 2.4.2).

2.4.1 M5 Flare Model Training with Default Data

Here, we show how to train the model with default data. To train the model with your own data:

1. You should first upload your file to the data directory (in the left hand side file list).
2. Edit the path to the training file:
   'train_data_file':'data/LSTM_M5_sample_run/normalized_training.csv'
   and replace the value 'data/LSTM_M5_sample_run/normalized_training.csv' with your new file name.

In [8]:
```
print('Loading the train_model function...')
from flarepredict_train import train_model
args = {'train_data_file':'data/LSTM_M5_sample_run/normalized_training.csv',
        'flare': 'M5',
       'modelid': 'custom_model_id'
      }
train_model(args)
```

```
Loading the train_model function...
turn off loggins is not supported
Starting training with a model with id: custom_model_id training data file: dat
a/LSTM_M5_sample_run/normalized_training.csv
Loading data set...
(84798, 10, 20)
Done loading data...
Training is in progress, please wait until it is done...

Finished training the M5 flare model, you may use the flarepredict_test.py prog
ram to make prediction.
```

2.4.2 Predicting with Your M5 Flare Model

To predict the testing data using the model you trained above, make sure the modelid value in the args variable in the following code is set exactly as the one used in the training, for example: 'custom_model_id'.

In [9]:
```
from flarepredict_test import test_model
args =  {'test_data_file': 'data/LSTM_M5_sample_run/normalized_testing.csv',
         'flare': 'M5',
         'modelid': 'custom_model_id'}
result_file_name = test_model(args)
```

```
turn off loggins is not supported
Starting testing with a model with id: custom_model_id testing data file: data/
LSTM_M5_sample_run/normalized_testing.csv
Loading data set...
(189, 10, 20)
Done loading data...
Formatting and mapping the data...
Prediction is in progress, please wait until it is done...
Finished the prediction task..
```

2.4.3 Reading the Results

The prediction result can be shown by passing the result variable to the pandas function read_csv as shown in the following example. The result shows the label your model predicted for each case.

In [10]:
```
import pandas as pd
df = pd.read_csv(result_file_name)
df
```

Out[10]:

	Predicted Label	label	flare	timestamp	NOAA	HARP	TOTUSJH	SAVNCPP	ABSNJZH	
0	Positive	Negative	N	2017-08-30T08:58:42.90Z	12673	7115	-0.498770	-0.503428	-0.374758	-0
1	Positive	Negative	N	2017-08-30T09:58:42.90Z	12673	7115	-0.503880	-0.510741	-0.418607	-0
2	Positive	Negative	N	2017-08-30T10:58:42.90Z	12673	7115	-0.488118	-0.504173	-0.449640	-0
3	Positive	Negative	N	2017-08-30T11:58:42.80Z	12673	7115	-0.481332	-0.500172	-0.444952	-0
4	Positive	Negative	N	2017-08-30T12:58:42.80Z	12673	7115	-0.473880	-0.501714	-0.419139	-0
...	...	...	...	...	...	...	...	...	...	
184	Negative	Negative	M1.3	2017-09-08T01:58:41.80Z	12673	7115	4.957265	3.600708	5.360394	3
185	Negative	Negative	M1.2	2017-09-08T02:58:41.80Z	12673	7115	4.998832	3.638885	5.449238	3
186	Negative	Positive	M8.1	2017-09-08T03:58:41.80Z	12673	7115	4.875287	3.552344	5.284544	3
187	Negative	Positive	M8.1	2017-09-08T04:58:41.80Z	12673	7115	4.843132	3.583137	5.129436	3
188	Negative	Positive	M8.1	2017-09-08T05:58:41.80Z	12673	7115	4.762235	3.627277	5.090359	3

189 rows × 26 columns

3. Acknowledgment

We thank the team of SDO/HMI for producing vector magnetic data products. The flare catalogs were prepared by and made available through NOAA NCEI. This work was supported by U.S. NSF grants AGS-1927578 and AGS-1954737.

4. References

DeepSun: Predicting Solar Flares Using a Long Short-term Memory Network

Hao Liu, Chang Liu, Jason T. L. Wang and Haimin Wang

https://iopscience.iop.org/article/10.3847/1538-4357/ab1b3c (https://iopscience.iop.org/article/10.3847/1538-4357/ab1b3c)

https://github.com/deepsuncode/LSTM-flare-prediction (https://github.com/deepsuncode/LSTM-flare-prediction)