Solar Flare forecasting from MF properties generated by SMART

Katarina Domijan

Maynooth University

January 28, 2015
Data: magnetic feature properties (MF) generated by Solar Monitor Active Region Tracker - SMART (Higgins et al. 2010).
• Data: magnetic feature properties (MF) generated by Solar Monitor Active Region Tracker - SMART (Higgins et al. 2010).
• Each MF detection: consider 20 most useful properties.
Data: magnetic feature properties (MF) generated by Solar Monitor Active Region Tracker - SMART (Higgins et al. 2010).

Each MF detection: consider 20 most useful properties.

For the purpose of this analysis all MF detections are treated as independent measurements.
Data: magnetic feature properties (MF) generated by Solar Monitor Active Region Tracker - SMART (Higgins et al. 2010).

- Each MF detection: consider 20 most useful properties.
- For the purpose of this analysis all MF detections are treated as independent measurements.
- Each MF detection is classified as flaring/non-flaring if it produced a flare within the next 24 hours following the observation.
Data: magnetic feature properties (MF) generated by Solar Monitor Active Region Tracker - SMART (Higgins et al. 2010).

Each MF detection: consider 20 most useful properties.

For the purpose of this analysis all MF detections are treated as independent measurements.

Each MF detection is classified as flaring/non-flaring if it produced a flare within the next 24 hours following the observation.

Aim: derive a classification rule to forecast the MF detections as flaring/non-flaring.
Data:

- I consider the data from 2003 to 2008 only.
Data:

- I consider the data from 2003 to 2008 only.
- In order to minimise the error caused by projection effects, only MF detections located within 30 deg from solar disc centre are considered in the analysis.
Data:

- I consider the data from 2003 to 2008 only.
- In order to minimise the error caused by projection effects, only MF detections located within 30 deg from solar disc centre are considered in the analysis.
- There are 3945 flaring and 102,876 non-flaring detections.
Data:

- I consider the data from 2003 to 2008 only.
- In order to minimise the error caused by projection effects, only MF detections located within 30 deg from solar disc centre are considered in the analysis.
- There are 3945 flaring and 102,876 non-flaring detections.
- MF properties are centered and scaled.
Training and testing data sets:

- **Training set**: randomly selects 300 instances of non-flares and 100 instances of flares from 2003 data.
- **Testing set**: all MF detections from 2003 excluding the training set.
- **Validation set**: all MF detections from 2004 to 2008.

We draw 500 random splits of training/testing data from the 2003 events. This can be useful for sensitivity analysis.
Training and testing data sets:

- **Training set**: randomly selects 300 instances of non-flares and 100 instances of flares from 2003 data.
- **Testing set**: all MF detections from 2003 excluding the training set.
**Training and testing data sets:**

- **Training set:** randomly selects 300 instances of non-flares and 100 instances of flares from 2003 data.
- **Testing set:** all MF detections from 2003 excluding the training set.
- **Validation set:** all MF detections from 2004 to 2008.

We draw **500 random splits** of training/testing data from the 2003 events. This can be useful for sensitivity analysis.
Training and testing data sets:

- **Training set**: randomly selects 300 instances of non-flares and 100 instances of flares from 2003 data.

- **Testing set**: all MF detections from 2003 excluding the training set.

- We draw **500 random splits** of training/testing data from the 2003 events. This can be useful for sensitivity analysis.

- **Validation set**: all MF detections from 2004 to 2008.
Feature selection:

- The **marginal relevance** score for each feature is derived from the training data set.

\[
\begin{align*}
&\text{MR score} \\
&R_{\text{val}}_{\text{Mx}} \\
&R_{\text{Str}} \\
&\text{WL}_{\text{sg Str}} \\
&\text{WL}_{\text{sg GpMm}} \\
&L_{\text{sg Mm}} \\
&L_{\text{nl Mm}} \\
&B_{\text{flux Mx}} \\
&B_{\text{fluxp Mx}} \\
&B_{\text{fluxn Mx}} \\
&M_{\text{xGrad GpMm}} \\
&M_{\text{MeanGrad}} \\
&M_{\text{MednGrad}} \\
&B_{\max G} \\
&B_{\min G} \\
&A_{\text{rea Mmsq}} \\
&B_{\text{mean G}} \\
&B_{\text{fluximb}} \\
&H_{\text{Glon wdth}} \\
&H_{\text{Gl lat wdth}} \\
&D_{\text{BfluxDt Mx}}
\end{align*}
\]
Feature selection:

- The **marginal relevance** score for each feature is derived from the training data set.
- We obtain 500 such scores for each feature (from the 500 training sets).
From the above plot marginal relevance score is consistent across random splits. From this we can choose overall best features.

- MxGradGpMm (Maximum gradient along polarity inversion line (PIL))
- BminG (largest negative magnetic field value)
- BmaxG (largest positive magnetic field value)
- BfluxnMx (negative magnetic flux)
- BfluxMx (total unsigned magnetic flux)
- AreaMmsq (area unit)
- MednGrad (median gradient along PIL)

Marginal relevance considers each feature separately. There is a correlation structure between features.
Feature selection:

- From the above plot marginal relevance score is consistent across random splits. From this we can choose overall best features.

- Features in order of their marginal relevance:
  1. MxGradGpMm (Maximum gradient along polarity inversion line (PIL))
  2. BminG (largest negative magnetic field value)
  3. BmaxG (largest positive magnetic field value)
  4. BfluxnMx (negative magnetic flux)
  5. BfluxMx (total un-signed magnetic flux)
  6. AreaMmsq (area unit)
  7. MednGrad (median gradient along PIL)
Feature selection:

- From the above plot marginal relevance score is consistent across random splits. From this we can choose overall best features.

- Features in order of their marginal relevance:
  1. MxGradGpMm (Maximum gradient along polarity inversion line (PIL))
  2. BminG (largest negative magnetic field value)
  3. BmaxG (largest positive magnetic field value)
  4. BfluxnMx (negative magnetic flux)
  5. BfluxMx (total un-signed magnetic flux)
  6. AreaMmsq (area unit)
  7. MednGrad (median gradient along PIL)

- Marginal relevance considers each feature separately. There is a correlation structure between features.
Feature selection:

- One training data set in 3 dimensions (3 features with the highest marginal relevance), red = flare, black = no flare.
Feature selection:

- One training data set in 2 dimensions, red = flare, black = no flare.
Feature selection:

- One testing data set in 3 dimensions (3 features with the highest marginal relevance), red = flare, black = no flare.
Feature selection:

- Validation set: in 3 dimensions (3 features with the highest marginal relevance), red = flare, black = no flare.
I ran a linear classifier (logistic regression) using the top 2 features selected by the marginal relevance algorithm.
I ran a linear classifier (logistic regression) using the top 2 features selected by the marginal relevance algorithm.

For each MF detection, the model outputs $p$, the probability of flare.
The classification rule from a training set can be applied to the validation set.

We can repeat this for 500 training sets.
Performance measures:

- I consider:
  - true positive rate (TPR) - flare predicted and observed,
  - false positive rate (FPR) - flare predicted but not observed,
  - true negative rate (TNR) - no flare predicted and none observed,
  - false negative rate (FNR) - no flare predicted but observed.

Overall forecast performance measure:

TSS - true skill statistic.
Performance measures:

- I consider:
  - true positive rate (TPR) - flare predicted and observed,
  - false positive rate (FPR) - flare predicted but not observed,
  - true negative rate (TNR) - no flare predicted and none observed,
  - false negative rate (FNR) - no flare predicted but observed.

- Overall forecast performance measure:
  - TSS - true skill statistic.
The classification rule from a training set was applied to a testing set for 500 random splits.

<table>
<thead>
<tr>
<th>p</th>
<th>TSS</th>
<th>TPR</th>
<th>TNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.84</td>
<td>0.97</td>
<td>0.87</td>
</tr>
<tr>
<td>0.2</td>
<td>0.85</td>
<td>0.95</td>
<td>0.90</td>
</tr>
<tr>
<td>0.3</td>
<td>0.85</td>
<td>0.93</td>
<td>0.92</td>
</tr>
<tr>
<td>0.4</td>
<td>0.83</td>
<td>0.89</td>
<td>0.94</td>
</tr>
<tr>
<td>0.5</td>
<td>0.80</td>
<td>0.85</td>
<td>0.95</td>
</tr>
<tr>
<td>0.6</td>
<td>0.75</td>
<td>0.79</td>
<td>0.96</td>
</tr>
<tr>
<td>0.7</td>
<td>0.68</td>
<td>0.71</td>
<td>0.97</td>
</tr>
<tr>
<td>0.8</td>
<td>0.59</td>
<td>0.61</td>
<td>0.98</td>
</tr>
<tr>
<td>0.9</td>
<td>0.44</td>
<td>0.45</td>
<td>0.99</td>
</tr>
</tbody>
</table>
The classification rule from a training set was applied to a testing set for 500 random splits. For each random split the TSS and the TPR and TNR were calculated by thresholding at $p \in (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)$. 

<table>
<thead>
<tr>
<th>$p$</th>
<th>TSS</th>
<th>TPR</th>
<th>TNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.84</td>
<td>0.97</td>
<td>0.87</td>
</tr>
<tr>
<td>0.2</td>
<td>0.85</td>
<td>0.95</td>
<td>0.90</td>
</tr>
<tr>
<td>0.3</td>
<td>0.85</td>
<td>0.93</td>
<td>0.92</td>
</tr>
<tr>
<td>0.4</td>
<td>0.83</td>
<td>0.89</td>
<td>0.94</td>
</tr>
<tr>
<td>0.5</td>
<td>0.80</td>
<td>0.85</td>
<td>0.95</td>
</tr>
<tr>
<td>0.6</td>
<td>0.75</td>
<td>0.79</td>
<td>0.96</td>
</tr>
<tr>
<td>0.7</td>
<td>0.68</td>
<td>0.71</td>
<td>0.97</td>
</tr>
<tr>
<td>0.8</td>
<td>0.59</td>
<td>0.61</td>
<td>0.98</td>
</tr>
<tr>
<td>0.9</td>
<td>0.44</td>
<td>0.45</td>
<td>0.99</td>
</tr>
</tbody>
</table>
Results: Testing sets

- The classification rule from a training set was applied to a testing set for 500 random splits.
- For each random split the TSS and the TPR and TNR were calculated by thresholding at $p \in (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)$.
- Below are median values for TSS, TPR and TNR over the 500 testing sets.

<table>
<thead>
<tr>
<th>p</th>
<th>TSS</th>
<th>TPR</th>
<th>TNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.84</td>
<td>0.97</td>
<td>0.87</td>
</tr>
<tr>
<td>0.2</td>
<td>0.85</td>
<td>0.95</td>
<td>0.90</td>
</tr>
<tr>
<td>0.3</td>
<td>0.85</td>
<td>0.93</td>
<td>0.92</td>
</tr>
<tr>
<td>0.4</td>
<td>0.83</td>
<td>0.89</td>
<td>0.94</td>
</tr>
<tr>
<td>0.5</td>
<td>0.80</td>
<td>0.85</td>
<td>0.95</td>
</tr>
<tr>
<td>0.6</td>
<td>0.75</td>
<td>0.79</td>
<td>0.96</td>
</tr>
<tr>
<td>0.7</td>
<td>0.68</td>
<td>0.71</td>
<td>0.97</td>
</tr>
<tr>
<td>0.8</td>
<td>0.59</td>
<td>0.61</td>
<td>0.98</td>
</tr>
<tr>
<td>0.9</td>
<td>0.44</td>
<td>0.45</td>
<td>0.99</td>
</tr>
</tbody>
</table>
The graphs below show boxplots of TSS, TPR and TNR at $p \in (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)$ from the 500 random splits.
The graphs below show median TSS, TPR and TNR at $p \in (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)$ as well as the 2.5th and 97.5th percentile.
The plots above can be used to compare different algorithms.
I considered logistic regression with $\geq 3$ features and classification algorithms with nonlinear classification rules.
For these training/testing sets I find that the linear model with only 2 features works as well.
The graphs below show median TSS, TPR and TNR at $p \in (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)$ as well as the 2.5th and 97.5th percentile.

<table>
<thead>
<tr>
<th>p</th>
<th>TSS</th>
<th>TPR</th>
<th>TNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.82</td>
<td>0.96</td>
<td>0.86</td>
</tr>
<tr>
<td>0.2</td>
<td>0.82</td>
<td>0.93</td>
<td>0.89</td>
</tr>
<tr>
<td>0.3</td>
<td>0.80</td>
<td>0.88</td>
<td>0.92</td>
</tr>
<tr>
<td>0.4</td>
<td>0.77</td>
<td>0.84</td>
<td>0.93</td>
</tr>
<tr>
<td>0.5</td>
<td>0.72</td>
<td>0.78</td>
<td>0.94</td>
</tr>
<tr>
<td>0.6</td>
<td>0.66</td>
<td>0.71</td>
<td>0.95</td>
</tr>
<tr>
<td>0.7</td>
<td>0.59</td>
<td>0.63</td>
<td>0.96</td>
</tr>
<tr>
<td>0.8</td>
<td>0.50</td>
<td>0.53</td>
<td>0.98</td>
</tr>
<tr>
<td>0.9</td>
<td>0.37</td>
<td>0.39</td>
<td>0.99</td>
</tr>
</tbody>
</table>
Results: Validation set

- Impact of random splits on predictions in 2004-2008
- Histogram of probabilities for two events (a non-flare and flare).

**MF detection No2**

<table>
<thead>
<tr>
<th>Frequency</th>
<th>0</th>
<th>40</th>
<th>80</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
</tr>
</tbody>
</table>

**MF detection No12**

<table>
<thead>
<tr>
<th>Frequency</th>
<th>0</th>
<th>50</th>
<th>150</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Results: Validation set 2004-2008

- False negatives:

![p of misclassified flares](chart)

- order in time

Katarina Domijan
ISSI Team meeting
Results: Validation set 2004-2008

- False negatives:
Results: Validation set 2004-2008

- False negatives:

```
p of misclassified flares
order by MxGradGpMm
```

Katarina Domijan  ISSI Team meeting
Conclusions

Simple model works well with small training sets. From the graphs of the data in 3-d it is clear that there is an overlap between the classes in this feature space - the classes are not perfectly separable. The shape of the classification boundary will not change this - there is limited scope for improvement in using more complex algorithms.

Results consistent across 500 training sets.

Katarina Domijan
ISSI Team meeting
Simple model works well with small training sets.
Conclusions

- Simple model works well with small training sets.
- From the graphs of the data in 3-d it is clear that there is an overlap between the classes in this feature space - the classes are not perfectly separable. The shape of the classification boundary will not change this - there is limited scope for improvement in using more complex algorithms.
Simple model works well with small training sets.

From the graphs of the data in 3-d it is clear that there is an overlap between the classes in this feature space - the classes are not perfectly separable. The shape of the classification boundary will not change this - there is limited scope for improvement in using more complex algorithms.

Results consistent across 500 training sets.
Future work

Use resampling for assessing probability of flare?

The individual measurements are not independent as some of the active regions are tracked through time. It might be beneficial to exploit this time structure.

Having time information could tell me how they behave in, say, 48 hours?

Getting the right features?
Future work

- Use resampling for assessing probability of flare?
Future work

- Use resampling for assessing probability of flare?
- The individual measurements are not independent as some of the active regions are tracked though time. It might be beneficial to exploit this time structure.
Future work

- Use resampling for assessing probability of flare?
- The individual measurements are not independent as some of the active regions are tracked though time. It might be beneficial to exploit this time structure.
- Having time information could tell me how they behave in, say, 48 hours?
Future work

- Use resampling for assessing probability of flare?
- The individual measurements are not independent as some of the active regions are tracked though time. It might be beneficial to exploit this time structure.
- Having time information could tell me how they behave in, say, 48 hours?
- Getting the right features?