Intelligence and statistics for rapid and robust earthquake detection, association and location

Anthony Lomax
ALomax Scientific, Mouans-Sartoux, France
anthony@alomax.net  www.alomax.net  @ALomaxNet

Alberto Michelini, Fabrizio Bernardi, and Valentino Lauciani
Istituto Nazionale di Geofisica e Vulcanologia, Roma, Italy
For effective earthquake and tsunami early-warning it is crucial that key earthquake parameters are determined as rapidly and reliably as possible.

**EarlyEst:** Rapid earthquake analysis module at INGV CAT tsunami alert center:

Realtime display OT+8min
Example: False events

Causes:
- Poor station distribution
- 3D structure but 1D velocity model
- Mis-picked phases
- Poor pick/travel-time error model
- ...

Use statistics and machine learning to identify problems.

Rapid, early results use minimal data: prone to bias & errors, poor magnitudes, false events, ...

False: M6 South Atlantic Ocean
False: M6 Mali
M3.5 Greece
M7 Mid-Atlantic Ridge
Identification of false events: apply statistics and machine learning to past true and (few) false events

“Data Frame” (2D array) of training data:

possible important attributes to discriminate true or false events

<table>
<thead>
<tr>
<th>true/false</th>
<th>ph_assoc</th>
<th>ph_used</th>
<th>dmin(deg)</th>
<th>gap2(deg)</th>
<th>err_oftime(sec)</th>
<th>lat(deg)</th>
<th>lon(deg)</th>
<th>depth(km)</th>
<th>mb</th>
<th>n</th>
<th>otime(UTC)</th>
<th>region</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>6</td>
<td>0.68</td>
<td>151.572</td>
<td>2.491</td>
<td>-13.31</td>
<td>132.195</td>
<td>10.16</td>
<td>-9</td>
<td></td>
<td></td>
<td>2015.05.01</td>
<td>Northern_Territory_Australia</td>
</tr>
<tr>
<td>21</td>
<td>17</td>
<td>11.024</td>
<td>135.796</td>
<td>2.805</td>
<td>-16.84</td>
<td>66.467</td>
<td>0.78</td>
<td>5.154</td>
<td>6</td>
<td></td>
<td>2015.04.19</td>
<td>Mid__Indian_Ridge</td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>1.771</td>
<td>166.38</td>
<td>2.639</td>
<td>-6.237</td>
<td>130.352</td>
<td>138.28</td>
<td>4.588</td>
<td></td>
<td></td>
<td>2015.06.17</td>
<td>Banda_Sea</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>0.636</td>
<td>247.126</td>
<td>1.508</td>
<td>38.496</td>
<td>21.876</td>
<td>10.16</td>
<td>-9</td>
<td></td>
<td></td>
<td>2015.06.17</td>
<td>Greece</td>
</tr>
<tr>
<td>45</td>
<td>39</td>
<td>30.86</td>
<td>75.262</td>
<td>5.01</td>
<td>-36.042</td>
<td>-16.471</td>
<td>7.81</td>
<td>5.573</td>
<td>26</td>
<td></td>
<td>2015.06.17</td>
<td>Southern_Mid__Atlantic_Ridge</td>
</tr>
<tr>
<td>44</td>
<td>38</td>
<td>3.688</td>
<td>98.833</td>
<td>1.073</td>
<td>2.598</td>
<td>124.161</td>
<td>332.42</td>
<td>4.907</td>
<td>31</td>
<td></td>
<td>2015.06.15</td>
<td>Celebes_Sea</td>
</tr>
<tr>
<td>65</td>
<td>56</td>
<td>1.702</td>
<td>90.896</td>
<td>1.419</td>
<td>27.306</td>
<td>140.285</td>
<td>383.2</td>
<td>4.892</td>
<td>48</td>
<td></td>
<td>2015.06.14</td>
<td>Bonin_Islands__Japan_Region</td>
</tr>
<tr>
<td>45</td>
<td>39</td>
<td>1.962</td>
<td>82.679</td>
<td>1.299</td>
<td>-0.232</td>
<td>124.559</td>
<td>66.41</td>
<td>5.041</td>
<td>32</td>
<td></td>
<td>2015.06.13</td>
<td>Southern_Molucca_Sea</td>
</tr>
<tr>
<td>36</td>
<td>32</td>
<td>1.678</td>
<td>125.296</td>
<td>1.572</td>
<td>45.355</td>
<td>142.046</td>
<td>283.2</td>
<td>4.623</td>
<td>22</td>
<td></td>
<td>2015.06.13</td>
<td>Hokkaido__Japan_Region</td>
</tr>
<tr>
<td>48</td>
<td>38</td>
<td>3.7</td>
<td>99.016</td>
<td>1.389</td>
<td>-3.27</td>
<td>143.81</td>
<td>10.16</td>
<td>5.166</td>
<td>31</td>
<td></td>
<td>2015.06.13</td>
<td>Near_N_Coast_of_New_Guinea_Png</td>
</tr>
<tr>
<td>37</td>
<td>26</td>
<td>14.836</td>
<td>128.997</td>
<td>1.308</td>
<td>13.704</td>
<td>120.71</td>
<td>250.78</td>
<td>4.843</td>
<td>21</td>
<td></td>
<td>2015.06.13</td>
<td>Mindoro__Philippines</td>
</tr>
<tr>
<td>78</td>
<td>65</td>
<td>5.116</td>
<td>121.641</td>
<td>3.102</td>
<td>-254.04</td>
<td>-175.774</td>
<td>19.53</td>
<td>5.133</td>
<td>37</td>
<td></td>
<td>2015.06.13</td>
<td>South_of_Tonga_Islands</td>
</tr>
<tr>
<td>34</td>
<td>34</td>
<td>0.699</td>
<td>113.06</td>
<td>1.23</td>
<td>34.446</td>
<td>137.382</td>
<td>322.6</td>
<td>4.604</td>
<td>18</td>
<td></td>
<td>2015.06.13</td>
<td>Near_S_Coast_of_Honshu__Japan</td>
</tr>
<tr>
<td>104</td>
<td>98</td>
<td>2.493</td>
<td>134.463</td>
<td>0.862</td>
<td>56.689</td>
<td>-156.725</td>
<td>34.77</td>
<td>5.21</td>
<td>84</td>
<td></td>
<td>2015.06.13</td>
<td>Alaska_Peninsula</td>
</tr>
<tr>
<td>53</td>
<td>50</td>
<td>1.731</td>
<td>125.087</td>
<td>1.592</td>
<td>38.951</td>
<td>142.607</td>
<td>34.77</td>
<td>5.027</td>
<td>36</td>
<td></td>
<td>2015.06.12</td>
<td>Near_East_Coast_of_Honshu__Japan</td>
</tr>
<tr>
<td>44</td>
<td>33</td>
<td>1.378</td>
<td>129.993</td>
<td>1.811</td>
<td>-5.562</td>
<td>152.304</td>
<td>19.53</td>
<td>5.138</td>
<td>31</td>
<td></td>
<td>2015.06.12</td>
<td>New_Britain_Region__P.N.G.</td>
</tr>
<tr>
<td>170</td>
<td>158</td>
<td>2.059</td>
<td>39.853</td>
<td>1.415</td>
<td>-15.56</td>
<td>-173.049</td>
<td>10.16</td>
<td>5.455</td>
<td>97</td>
<td></td>
<td>2015.06.12</td>
<td>Tonga_Islands</td>
</tr>
<tr>
<td>50</td>
<td>41</td>
<td>4.058</td>
<td>93.874</td>
<td>2.858</td>
<td>-20.805</td>
<td>-173.832</td>
<td>10.16</td>
<td>5.058</td>
<td>23</td>
<td></td>
<td>2015.06.12</td>
<td>Tonga_Islands</td>
</tr>
<tr>
<td>53</td>
<td>38</td>
<td>1.663</td>
<td>144.499</td>
<td>1.695</td>
<td>8.009</td>
<td>125.963</td>
<td>96.09</td>
<td>5.026</td>
<td>34</td>
<td></td>
<td>2015.06.11</td>
<td>Mindanao__Philippines</td>
</tr>
<tr>
<td>36</td>
<td>26</td>
<td>3.828</td>
<td>118.996</td>
<td>1.688</td>
<td>12.902</td>
<td>146.272</td>
<td>19.53</td>
<td>5.179</td>
<td>23</td>
<td></td>
<td>2015.06.11</td>
<td>South_of_Mariana_Islands</td>
</tr>
<tr>
<td>41</td>
<td>33</td>
<td>1.543</td>
<td>161.243</td>
<td>0.814</td>
<td>29.73</td>
<td>131.359</td>
<td>9.96</td>
<td>5.032</td>
<td>27</td>
<td></td>
<td>2015.06.11</td>
<td>Southeast_of_Ryukyu_Islands</td>
</tr>
<tr>
<td>83</td>
<td>74</td>
<td>1.981</td>
<td>133.718</td>
<td>1.902</td>
<td>38.677</td>
<td>143.213</td>
<td>19.53</td>
<td>5.284</td>
<td>67</td>
<td></td>
<td>2015.06.11</td>
<td>Off_East_Coast_of_Honshu__Japan</td>
</tr>
<tr>
<td>178</td>
<td>185</td>
<td>2.102</td>
<td>81.348</td>
<td>1.391</td>
<td>39.503</td>
<td>143.367</td>
<td>19.73</td>
<td>5.328</td>
<td>138</td>
<td></td>
<td>2015.06.11</td>
<td>Off_East_Coast_of_Honshu__Japan</td>
</tr>
</tbody>
</table>

“labelled” - identified as true or false event

Problem: many true and few false events!
First step: basic statistical & expert analysis of past true and (few) false events

Statistics: scatter matrix: examine pairs of attributes

- **Secondary gap outliers** → false events located by compact, distant clusters of stations
- **Outliers in origin-time error** → large pick residuals for false events
- **Depth outliers in longitude** → false events sometimes deep and in a-seismic regions
Second step: Machine learning (classification, regression, ...) for identifying outliers, making decisions, finding patterns...

Examine semi-automatically in high-dimension many attributes

What is machine learning?

Given training data, construct an algorithm to make predictions on new data.

1. Learn (select and tune algorithms) using training data.
2. Test algorithm on testing data.
3. Apply algorithm to new data.

• Supervised learning: predict attributes of data:
  - Classification: learn from labeled, xy training data how to predict the (discrete) class y of new, unlabeled data x.
  - Regression: learn from xy training data how to predict the (continuous) values of y variables in new data x.

• Unsupervised learning:
  No target attributes, try to discover clustering or distribution of the data, or reduce the dimensionality of the data.

Applications:

Decision / classification (e.g. False event? Tsunamigenic earthquake?)

Outlier detection (e.g. False event? Unusual event?)

and many more...

http://scikit-learn.org
Machine learning: multitude of methods depending on goals and characteristics of data set

classification
- SVC
- Ensemble Classifiers
- Naive Bayes
- Text Data
- Linear SVC
- kNeighbors Classifier
- SGD Classifier
- kernel approximation

clustering
- Spectral Clustering
- GMM
- KMeans
- MeanShift
- VBGMM
- MiniBatch KMeans

regression
- SGD Regressor
- Lasso
- ElasticNet
- Ridge Regression
- SVR(kernel=’rbf’)
- SVR(kernel=’linear’)

dimensionality reduction
- Randomized PCA
- Isomap
- Spectral Embedding
- LLE

Identify false events
- START
- get more data
- >50 samples
- few features should be important
- <100K samples

scikit-learn algorithm cheat-sheet
http://scikit-learn.org
Multiple machine learning algorithms: train and test with past true and (few) false events

Support vector machines (SVMs)

Classifier Algorithms:

1. SVC with linear kernel
2. LinearSVC (linear kernel)
3. SVC with RBF kernel
4. SVC with polynomial (degree 3) kernel

Nearest Neighbors Classification

Data Frame (2D array) of training data:
Possible important attributes to discriminate true or false events

“labelled” true or false event

Problem: many true and few false events!
### Multiple machine learning algorithms: train and test with past true and (few) false events

#### Algorithms act in high-dimension using many attributes
→ may discover complex relationships between attributes,
→ may be difficult to understand in terms of expert knowledge & scientific theory.

#### Many algorithms to select and parameters to tune
→ great open software helps.
Intelligence and statistics for rapid and robust earthquake analysis, identification of false events: Conclusions

- **Statistical analysis** aids in acting on individual or few attributes, (e.g. stronger filtering on azimuth gaps)
  Direct use of expert knowledge & scientific theory

- **Machine learning** acts in high-dimension using many attributes:
  Powerful and shows much promise for improving early warning reliability,
  Many machine learning algorithms are very familiar in geophysics, and powerful new algorithms for big data, image recognition, …
  BUT, automated, not theory based

- False events: Include past event history? → Recursive Neural Networks?

- Easy to use with well documented, open tools in Python, R, Java, …

- What advantages and trade-offs for science?
  Machine learning, Automation ↔ Expert knowledge, Scientific Theory

Support: Centro Nazionale Terremoti, INGV
Data: ingv.it, geofon.gfz-potsdam.de, geosbud.ipgp.fr, resif.fr, ird.nc, iris.washington.edu, usgs.gov
Software: Python: scikit-learn.org, pandas.pydata.org, matplotlib.org; R statistics language