

The Simpler the Better

Human-readable models

V. Lakshmanan^{1,2}

¹Cooperative Institute of Mesoscale Meteorological Studies
University of Oklahoma

²Radar Research and Development Division
National Severe Storms Laboratory

January 6, 2009

Built three different AI models to predict the category (frozen, liquid or none) based on polarimetric radar variables:

- long-shot model for the purpose of winning the competition (it didn't).
- a complex black box model to represent the average learnability of the dataset.
- simple, human-readable model that would perform similarly to the above two models but possess the additional advantage of being easy to implement and comprehend.

The number of training (847) and testing (346) patterns misleadingly large.

- Dataset was concentrated in Oklahoma
- Collected on a handful of winter events
- True number of independent training and testing instances far less.

There should be no virtual difference in skill between simple and more complex models on such a small dataset.

- Any good approach would be middle-of-the-pack.
- The most easily justifiable approach is to build simple models.
- **If** the testing patterns were chosen from the same set of cases, then it was likely that a nearest neighbor approach might even beat out the competition even if it would not be operationally feasible.

Built three AI models:

- a nearest neighbor approach because we felt this afforded a reasonable chance of actually winning
- a neural network
- a simple decision tree

Result:

- The nearest neighbor approach did not win because the creator of the data set had the foresight to choose the test patterns from a different pool than the training patterns!
- But as expected, the NN and decision tree were well within the performance bounds of the best submitted entries.
- The decision tree is simple enough to fit on a slide

If a human-readable data-driven model provides results statistically similar to complex methods, then the non-technical benefits of a human-readable model should cause it to be selected.

It would be a better candidate for operational implementation than techniques that are harder to comprehend or implement.

Omitted attributes that should, physically, not matter in the final classification, using only the following attributes:

- 1 Temperature in Celsius
- 2 Relative Humidity
- 3 Speed (computed from the u and v components)
- 4 Freezing level
- 5 Height above freezing (computed by subtracting the height from the freezing level)
- 6 Zdr
- 7 RhoHV
- 8 Kdp
- 9 Z

Created two new features that form a pre-classification

- whether the pattern in question corresponds to "clear air" or to "frozen"
- Divided up the dataset into two parts: those for which the category was "none" and the rest (similarly for frozen/not-frozen)
- Trained a Quinlan's C 4.5 decision tree with heavy pruning and confidence measures

CLEARAIR=1 if:

```
RhoHV <= 0.768208 AND  
Z <= 10.626065 AND  
RelH > 81 AND  
( HtAboveFreezing <= 853 OR  
HtAboveFreezing > 1113)
```

Otherwise, it was set to zero.

FROZEN=0 if:

```
RhoHV <= 0.487228 OR (TmpC > -2.143738 AND  
Z <= 24.4076)
```

Otherwise, it was set to one.

Used Weka and the following parameters:

- 1 Temperature in Celsius
- 2 Relative Humidity
- 3 Speed (computed from the u and v components)
- 4 Freezing level
- 5 Height above freezing (computed by subtracting the height from the freezing level)
- 6 Zdr
- 7 RhoHV
- 8 Kdp
- 9 Z
- 10 CLEARAIR
- 11 FROZEN

The training of each of the models followed a cross-validation approach. The model was trained on 90% of the data and tested on 10%. This was repeated ten times and the average cross-validated accuracy (percent correct) was used as the measure of the "goodness" of the model.

First model we trained was a nearest-neighbor model.

- A candidate pattern was assigned to a category based on the category of its 10 closest neighbors in the training set.
- Distance measure? Straight Euclidean distance of the 11 feature attributes!
- Weighted the vote of each neighbor by the reciprocal of its Euclidean distance.

Why is K-NN a long-shot to win?

- The cross-validation samples (the 10%) were selected randomly from the training dataset
- Likely that at least some of the patterns were very similar.
- The cross validation accuracy of the nearest-neighbor approach was the highest of the three approaches we tried – 72%.
- **If** the test patterns were similarly chosen from the same pool, this rather simplistic approach even had a chance of winning.
- So, this was our official entry.

The next two approaches were aimed towards being in the middle of the leaders, not necessarily win.

However, the goal of these two approaches was simplicity.

The neural network chosen for simplicity of implementation and retraining

- An operational implementation can simply read the new weights from a file
- Trained by backpropagation with a learning rate of 0.1, momentum of 0.2 with 45% of the training samples (i.e. 45% of the 90% used for training) used for early stopping. There were 7 hidden nodes.
- The neural network had a cross-validation accuracy of 68%.

The decision tree chosen for human readability.

- Relatively simple to examine
- Humans often want to develop a "feel for the model"
- Better if model is a simple, easy to understand set of rules.

Quinlan's C4.5 decision with a pruning threshold of 0.4 and stopping the splitting of nodes if it would result in leaves with less than 20 instances. This simple model had a cross-validation accuracy of 68%.

```
clearair <= 0
| frozen <= 0
| | tmpc <= -3.549988: frozen (35.0/15.0)
| | tmpc > -3.549988: liquid (168.0/64.0)
| frozen > 0
| | speed <= 1.431182: liquid (47.0/19.0)
| | speed > 1.431182: frozen (436.0/83.0)
clearair > 0
| tmpc <= 1.575012
| | Z <= -25.6025: none (31.0/8.0)
| | Z > -25.6025
| | | frzl <= 853: none (30.0/9.0)
| | | frzl > 853
| | | | tmpc <= -7.324982: none (29.0/14.0)
| | | | tmpc > -7.324982: frozen (51.0/19.0)
| tmpc > 1.575012: liquid (20.0/9.0)
```

Obviously, true "clear-air" cases are needed in warmer

- test data set was chosen from a different pool of instances than the training data set.
- Not surprisingly, the nearest-neighbor approach did not win. Its True Skill Score (TSS) on the test data set was 0.27.
- As expected, the methods all came in the middle of the pack, statistically not different from the other entries.
- The neural network approach had a TSS of 0.30 and the decision tree approach had a TSS of 0.28.

If the other entries are, as we expect, considerably more complex, we suggest the operational use of the decision tree above. **The complete human-readable description of our rules fits on half-a-page.**

Funding for this research was provided under NOAA-OU Cooperative Agreement NA17RJ1227. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of the National Severe Storms Laboratory (NSSL) or the U.S. Department of Commerce.