The 2008 Artificial Intelligence Competition

Valliappa Lakshmanan
National Severe Storms Laboratory and University of Oklahoma

Elizabeth E. Ebert
Bureau of Meteorology Research Center, Australia

Sue Ellen Haupt
Pennsylvania State University

Abstract

The Artificial Intelligence Committee of the AMS conducted a classification competition to correctly classify storms into one of four types based on radar-derived characteristics. This competition, which was sponsored by Weather Decision Technologies, received seven entries, three of which beat the baseline decision tree used in the original work that used the dataset.

We describe the dataset for automated storm type classification. The dataset is the result of a clustering program run on multi-radar reflectivity images and properties of those clusters computed by a set of severe weather algorithms. The labels on the dataset, created by manual identification, identify each cluster as being one of four categories: supercells, convective lines, pulse storms or non-organized cells. The ranking was carried out using the True Skill Statistic.

Introduction

The Artificial Intelligence Comittee is one of the technical advisory committees of the American Meteorological Society. Its mission is to ensure that members of the Society are informed about and encouraged to use modern artificial intelligence techniques that can contribute meaningfully to their scientific research and algorithm development activities. The term artificial intelligence denotes a large body of advanced computer techniques for data analysis, knowledge discovery and operational system development that have been productively applied to a large number of scientific and industrial applications.

The AI committee of the AMS organizes scientific conferences and has organized (typically every other year) a tutorial session before the AMS annual meeting. These tutorial sessions have been very popular and have served to introduce many AMS members to AI techniques. However, they have gotten very repetitive the past few years, since the same committee members ended up presenting variants of the same tutorials over the years.

In a bid to increase the variety of techniques covered, and to introduce the various techniques possible when analyzing the same dataset, we decided to conduct an AI competition. We approached probability and statistics experts at the AMS, and Beth

Ebert agreed to help us evaluate the entries. Weather Decision Technologies of Norman, OK kindly agreed to sponsor the competition by underwriting the prizes to the winners.

The purpose of the AI competition, then, is to provide a common dataset to showcase a variety of machine intelligence techniques. The competitive aspect is just a side-effect, and is not intended to be a primary focus.

Dataset

The dataset used for the 2008 competition was created to address two separate research issues. The immediate reason was to answer the question of whether the skill score of a forecast office as evaluated by the National Weather Service depended (to a statistically significant extent) on the type of storms that the forecast office faced that year. In other words, was it more the impact of the type of weather than it was of forecaster skill? It is not possible to manually identify the type of storms in the amount of radar imagery required to answer this question. Therefore, an automated storm-type identification technique was desired. The longer term reason to create the dataset and implement a machine intelligence technique is to create a national storm events database that can be used to support spatiotemporal queries based on storm attributes and geographic information. For this too, an automated storm and storm type identification technique was required.

Eric Guillot et. al (2008) present the procedure for creating the dataset, which is summarized here. The raw data comes from the CONUS WDSS-II system (described in Lakshmanan et. al 2007) and consists of data combined in real-time from more than 130 WSR-88D Doppler radars. The inputs are gridded fields like the reflectivity composite, VIL, echo top heights, hail diagnostic products, low-level and mid-level shear.

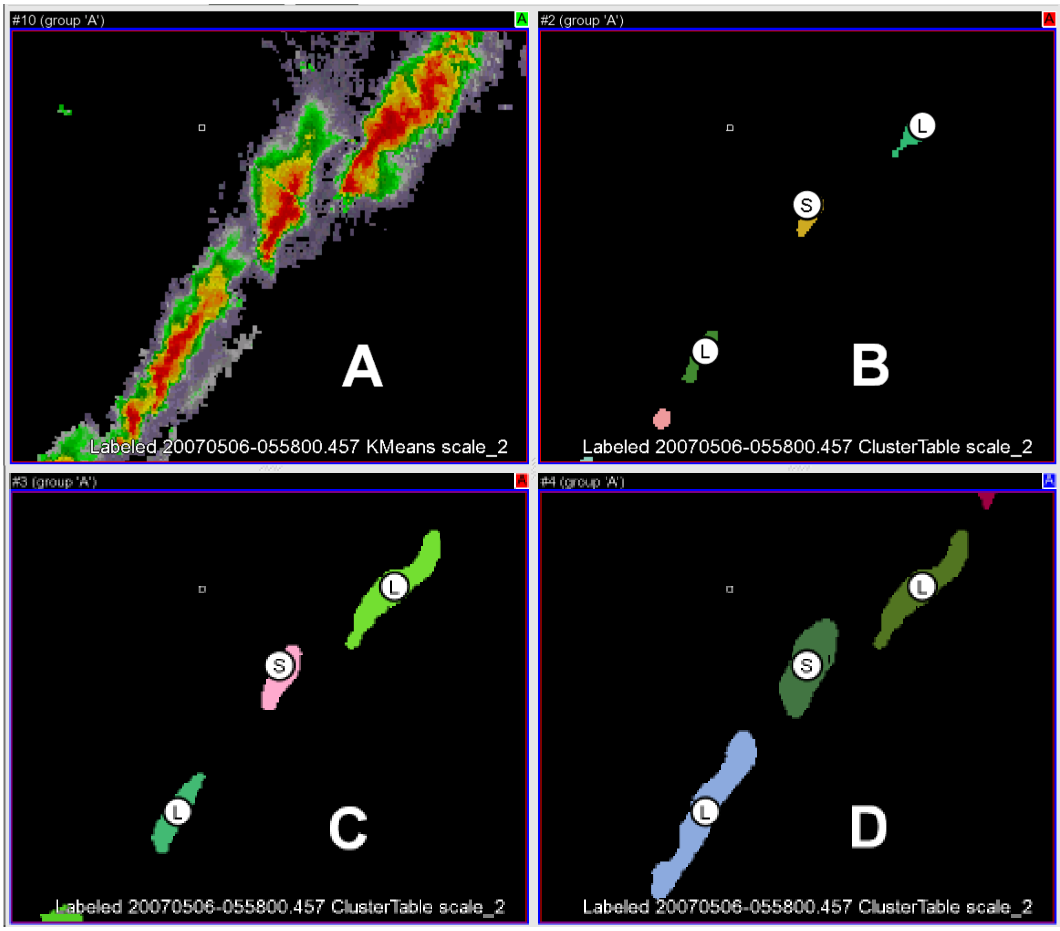

Figure 1: Clustering radar reflectivity to identify storms at different scales.
Using a hierarchical clustering technique (Lakshmanan et. al 2003), reflectivity composite images combined from multiple radars were clustered to identify storms at different scales. The largest scale (shown in the panel marked D, and at 420 km^2 scale) was used to extract storm attributes.

Guillot et. al (2008) manually classified over 1,000 storms over three days worth of data (March 28th, May 5th, and May 28th of 2007). The human expert used all the fields available to the automated algorithm to draw geographic polygons. Each polygon (see the second panel in the first row of Figure 2) was tagged as having storms in one of four categories: non-organized, supercells, linear convective storms or pulse storms.

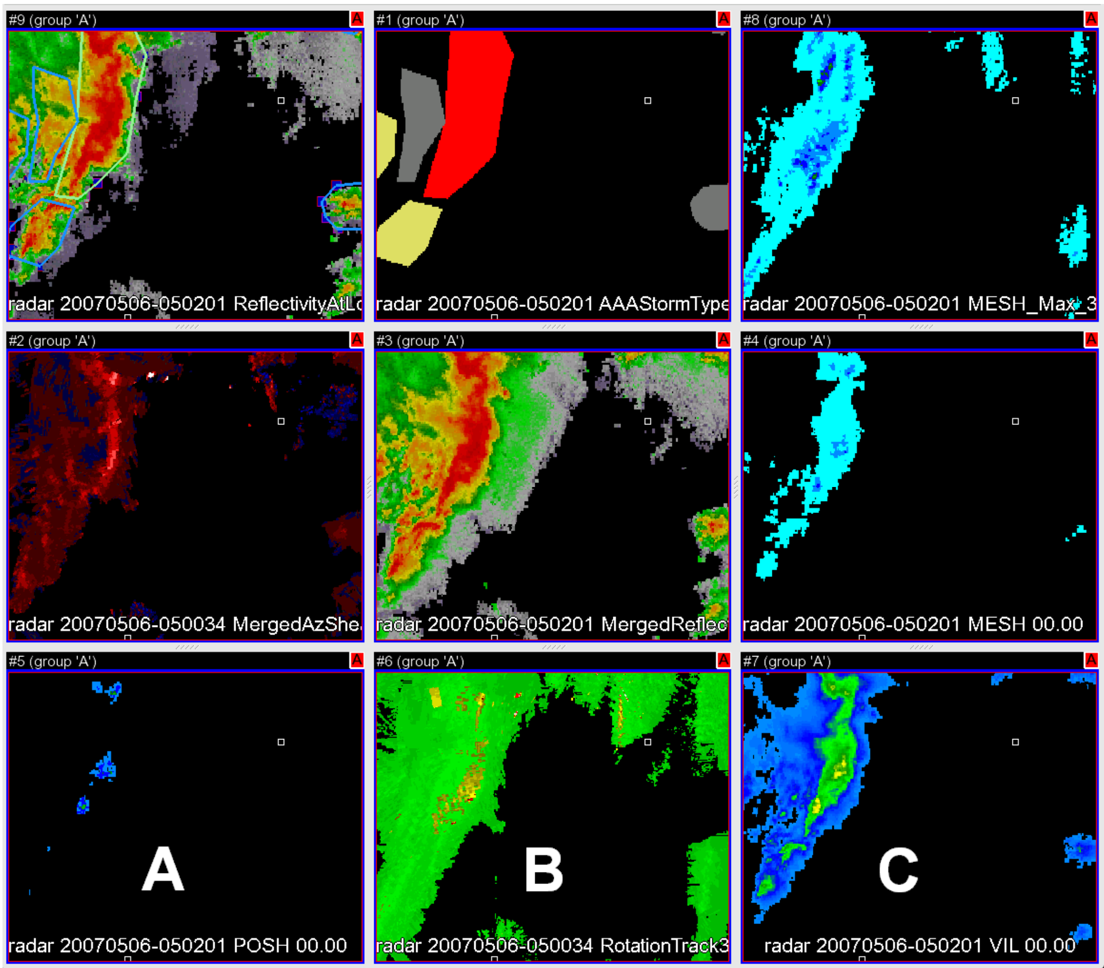

Figure 2: The process of manual classification by drawing geographic polygons.

All the clusters identified within each polygon were then classified as being of the particular storm type. The following attributes were extract for each cluster based on the polygons (a few of the relevant gridded fields are shown in Figure 2):

AspectRatio	dimensionless	An ellipse is fitted to the storm. This is the ratio of the length of the major axis to the length of the minor axis of the fitted ellipse.
ConvectiveArea	km^2	Area of the storm that is convective
LatRadius	km	Extent of the storm in the north-south direction
LatitudeOfCentroid	Degrees	Location of storm's centroid
LifetimeMESH	mm	Maximum expected hail size of the storm over its entire past history

LifetimePOSH	dimensionless	Peak probability of severe hail of the storm over its entire past history
LonRadius	km	Extent of the storm in the east-west direction
LonRadius	km	Extent of the storm in the east-west direction
LongitudeOfCentroid	Degrees	Location of the storm's centroid
LowLvlShear	s^-1	Shear closest to the ground as measured by radar
MESH	mm	Maximum expected hail size from storm
MaxRef	dBZ	Maximum reflectivity observed in storm
MaxVIL	kg/m^2	Maximum vertical integrated liquid in storm
MeanRef	dBZ	Mean reflectivity within storm
MotionEast	MetersPerSecond	Speed of storm in easterly direction
MotionSouth	MetersPerSecond	Speed of storm in southerly direction
OrientationToDueNorth	degrees	Orientation of major axis of ellipse to due north. A value of 90 indicates a storm that is oriented east-west. The more circular a storm is (see aspect ratio), the less reliable this measure is.
POSH	dimensionless	Peak probability of severe hail in storm
Rot120	s^-1	Peak probability of severe hail in storm
Rot30	s^-1	Maximum azimuthal shear observed in storm over the past 30 minutes
RowName	dimensionless	Storm id
Size	km^2	Storm size
Speed	MetersPerSecond	Speed of storm

Table 1: The parameters used in the decision tree

Guillot et. al (2008) then trained a decision tree (Quinlan 1993) to classify storms into one of the four types in an automated manner. They then used the decision tree to classify several days of data in order to answer the original question about forecaster skill. Part of the decision tree used by Guillot et. al is shown in Figure 3.

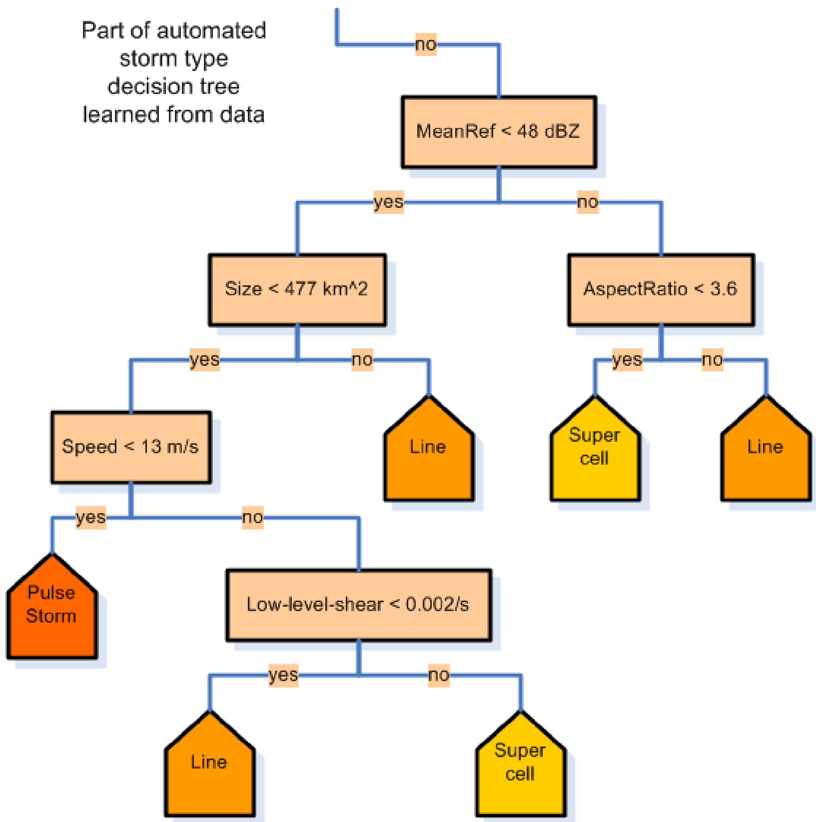

Figure 3: Part of the automated storm type decision tree used by Guillot et. al (2008)

On an independent dataset of 3 days, the decision tree was found to have a True Skill Statistic, defined as

$$HK = \frac{hits}{hits + misses} - \frac{false\ alarms}{false\ alarms + correct\ \ negatives}$$

of 0.58. The aim of the competition was to develop a better storm type classifier.

Results

We received six official, and one unofficial, entry to the competition. The unofficial entry, by Neil Gordon, was not accompanied by an AMS manuscript. So, even though it is included in comparisons, it is not eligible for a prize (it would have placed third, just a hair behind the eventual winners, if it had been eligible).

The official entries were:

1. John K. Williams and Jenny Abernathy who used random forests and fuzzy logic
2. Ron Holmes, who used a neural network

3. David Gagne and Amy McGovern who used boosted decision tree
4. Jenny Abernathy and John Williams who used support vector machines
5. Luna Rodriguez who used genetic algorithms
6. Kimberly Elmore who used discriminant analysis and support vector machines

An examination of the frequency with which different techniques predicted the different storm categories points out that the Holmes entry gets the distribution of Category 0 (the common case) wrong whereas the Rodriguez entry essentially classifies all storms as Category 0. The other entries are all in the ball park in terms of getting the frequency correct.

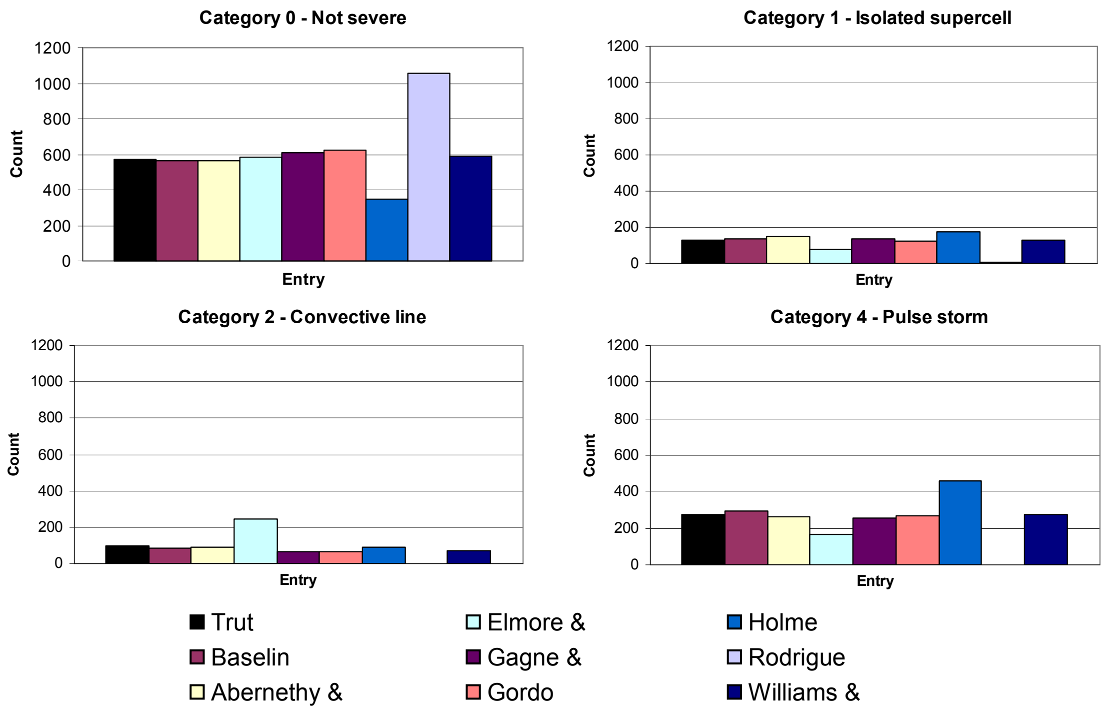

Figure 4: Frequency with which different categories were predicted by the various techniques

Several of the entered machines were very similar in the way they classified the instances of the test data set. It is apparent that the Gagne, Gordon and Williams entries are very similar.

	Truth	Baseline	Abernethy & Williams	Elmore & Richman	Gagne & McGovern	Gordon	Holmes	Rodriguez	Williams & Abernethy
Truth	**100**	**74**	**72**	**67**	**77**	**76**	**62**	**53**	**77**
Baseline	**74**	**100**	**77**	**69**	**84**	**84**	**62**	**52**	**84**
Abernethy & Williams	**72**	**77**	**100**	**70**	**83**	**80**	**61**	**52**	**83**
Elmore & Richman	**67**	**69**	**70**	**100**	**75**	**76**	**54**	**55**	**73**
Gagne & McGovern	**77**	**84**	**83**	**75**	**100**	**93**	**62**	**57**	**93**

Gordon	**76**	**84**	**80**	**76**	**93**	**100**	**61**	**58**	**91**
Holmes	**62**	**62**	**61**	**54**	**62**	**61**	**100**	**32**	**62**
Rodriguez	**53**	**52**	**52**	**55**	**57**	**58**	**32**	**100**	**55**
Williams & Abernethy	**77**	**84**	**83**	**73**	**93**	**91**	**62**	**55**	**100**

Figure 5: Similarity of the machines developed by the teams in terms of their classification of the test dataset.

The True Skill Statistic of the various entries is shown below:

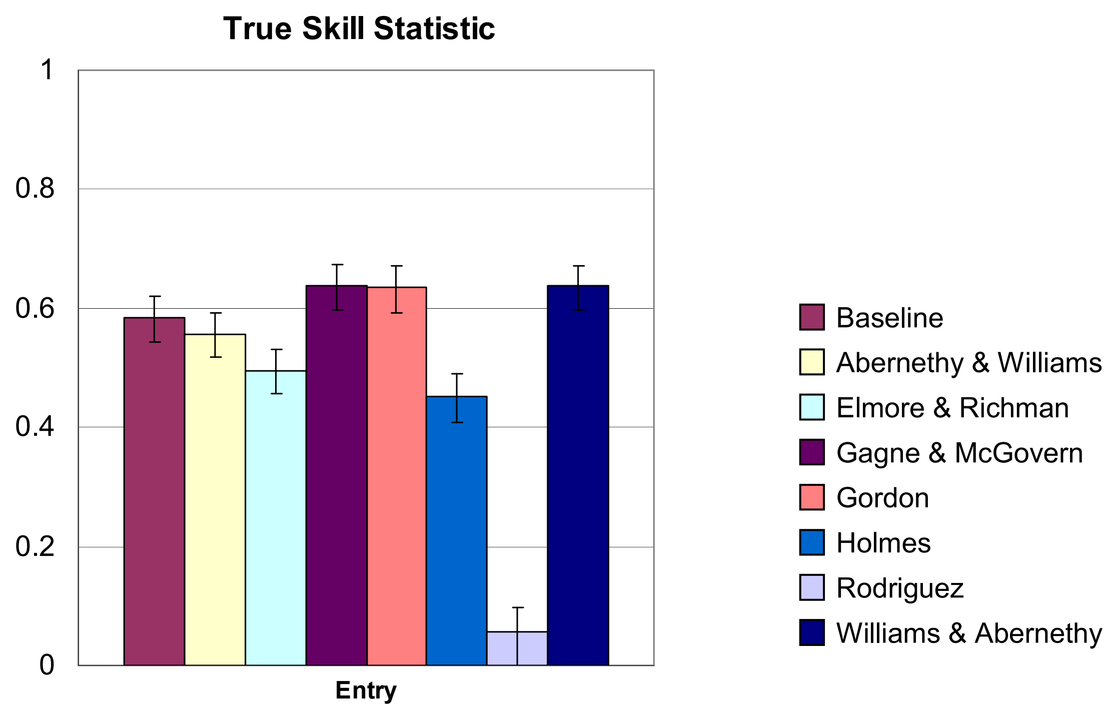

Figure 6: True Skill Statistic of the entries on the test dataset.

Conclusion

Based on the above results, the entries by Gagne & McGovern and Williams & Abernathy were judged to be joint-first. The entry from Abernathy & Williams was judged to be third.

Conclusion

We thank Weather Decision Technologies for their sponsorship of this competition and all the participants for entering the competition and explaining their methodology in the papers that accompany this session. We understand how hard it can be to carve out time to carry out extra-curricular tasks and we are grateful to all for participating.

References

Lakshmanan, V., R. Rabin, and V. DeBrunner, 2003: Multiscale storm identification and forecast. *J. Atm. Res.*, **67**, 367-380

Guillot, E., T.M.Smith, V. Lakshmanan, K. L. Elmore, D.W. Burgess and G.J. Stumpf, 2008: Tornado and Severe Thunderstorm Warning Forecast Skill and its Relationship to Storm Type, Applications in Meteorology, Oceanography, Hydrology and Climatology, New Orleans. CD-ROM 4 A.3

Lakshmanan, V., T. Smith, G. J. Stumpf, and K. Hondl, 2007c: The warning decision support system - integrated information. *Weather and Forecasting*, **22**, 596-612

Quinlan, J. R. 1993. C4.5: Programs For Machine Learning. Morgan Kaufmann, Los Altos.