Automatic Thunderstorm Detection via Boosting Using LM Output (Master Thesis – Preliminary results)

Donat Perler, Oliver Marchand

MeteoSchweiz, Bereich Wetter, Prozess Modelle
ETH Zürich, Departement Informatik, Institute for Computational Science

2006-03-07

Overview

Overview

Data Sources and Work Flow

Learn Category and Data Sources

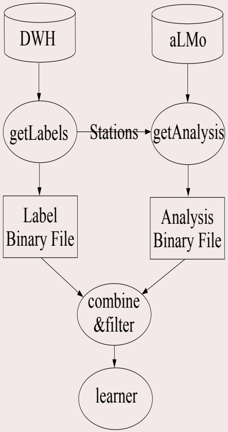
Learn category: Supervised Learning

Data: 127622 tuples $(\mathbf{x}^{(i)}, y^{(i)})$
one year data
over Switzerland

$y^{(i)}$ Synop- and lightning-messages from 80 Swiss weather stations

$\mathbf{x}^{(i)}$ Hourly aLMo analysis (nearest grid point to a station $\cup$ 8 neighbor points) with spacial resolution of 7 km

Work Flow



Data Sources and Work Flow

Learn Category and Data Sources

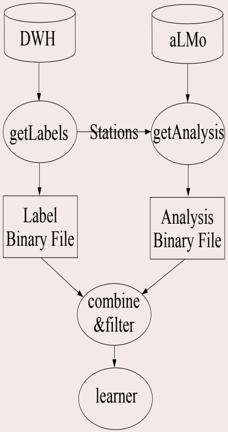
Learn category: Supervised Learning

Data: 127622 tuples $(\mathbf{x}^{(i)}, y^{(i)})$
one year data
over Switzerland

$y^{(i)}$ Synop- and lightning-messages from 80 Swiss weather stations

$\mathbf{x}^{(i)}$ Hourly aLMo analysis (nearest grid point to a station $\cup$ 8 neighbor points) with spacial resolution of 7 km

Work Flow



Constitution of the Feature Space and the Classes

Classes

Classes: $\mathcal{Y} = \{\text{thunderstorm, no thunderstorm}\}$

Thunderstorm: ww-codes $\{13, 17, 29, 91 \ldots 99\}$ or lightning detection data $(\#((\Delta t < 0.5\text{h}) \wedge (\Delta x < 13.5\text{km})) \geq 3)$

No thunderstorm: otherwise

Choice of the Features

Feature Space $\mathcal{X}$

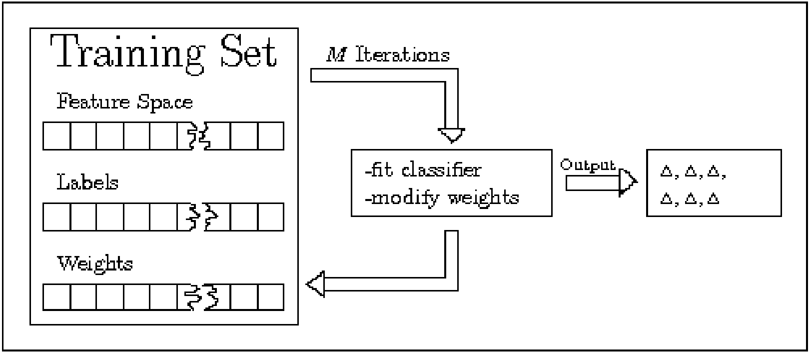
The feature space consists of 51 **features** which were turned out to be relevant for thunderstorm prediction (see Neumann 1971 [1], Huntrieser 1995 [2], Ducrocq 1998 [4], Calas 2000 [5], Haklander 2003 [6]). Most of the features are based on **average values** from 3×3 profiles of the aLMo analysis.

The features are:

- aLMo-model variables (T, QV, wind direction and velocity, etc. on 4 different reference levels; pressure, etc.)
- Temporal and spatial information (time of day, etc.)
- Composed variables (Θ_e, wind shear, KOI, CAPE, SWEAT, LI, surface moisture flux convergence, etc.)

AdaBoost: How does learning work?

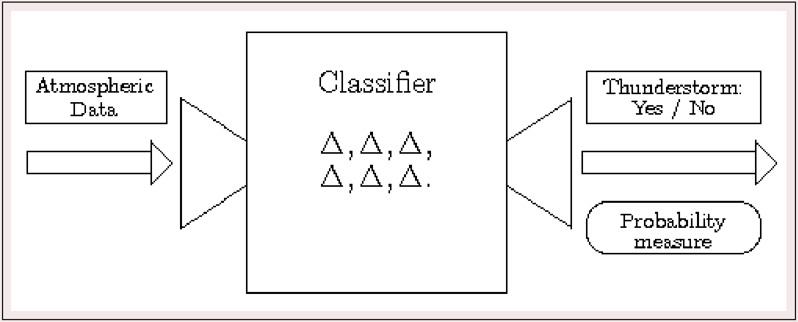
Learning Phase



Introduced by Freund and Shapire 1997 [3]

AdaBoost: Application

Application of the learn Classifier



Overview

1. Automatic Weather Interpretation
   - Data Sources
   - Definition of the Feature Space and the Classes
   - Boosting-Algorithms
2. Results
3. Conclusion

Results

Table showing skill scores for different types of classifiers

Classifier	POD	FAR	FBI	CSI	HSS
DWD	18%	94%	3.12	0.05	0.08
DWD opt. for CH	45%	68%	1.42	0.23	0.34
AdaBoost	57%	59%	1.44	0.32	0.46
AdaBoost & EFS	72%	34%	1.10	0.52	0.67

Explanation:

AdaBoost AdaBoost using decision stumps

EFS Enhanced feature space (with 51 features)

Results

Table showing skill scores for different types of classifiers

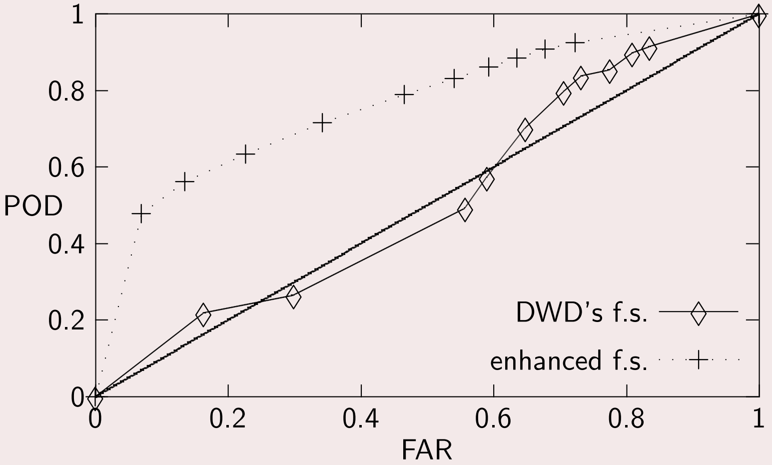
Classifier	POD	FAR	FBI	CSI	HSS
DWD	18%	94%	3.12	0.05	0.08
DWD opt. for CH	45%	68%	1.42	0.23	0.34
AdaBoost	57%	59%	1.44	0.32	0.46
AdaBoost & EFS	72%	34%	1.10	0.52	0.67

Explanation:

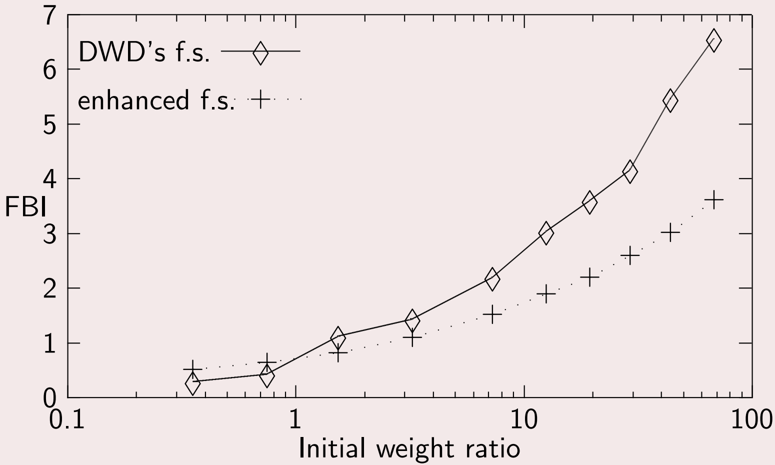
AdaBoost AdaBoost using decision stumps

EFS Enhanced feature space (with 51 features)

Influence of the expansion of the feature space looking at the ROC-curves of AdaBoost applied to both spaces



Change of the initial weight ratio



Overview

1. Automatic Weather Interpretation
   - Data Sources
   - Definition of the Feature Space and the Classes
   - Boosting-Algorithms
2. Results
3. Conclusion

Summary

Advantages of Boosting

- Probability measure as additional information
- Improvement by the use of AdaBoost
- Improvement by the expansion of the feature space
- Interpretability of the results
- Short learning time ($t < 10$ minutes)

Drawbacks of Boosting

- Learning process must be performed
- Data must be available

Further improvement and future work

Improvments

Data:

- Radar data as additional data source
- Longer time period for training (for example 5 years)
- Expansion of the feature space by quantities which give information about temporal changes.

Algorithm:

- testing of further base classifiers
- Regression instead of classification

Further improvement and future work

Future Work

- generating graphical output
- make it ready for operational use
- Classification of further weather types

Thank you for your attention

- Questions
- Discussion

- [Neumann, 1971] C. J. Neumann
  The thunderstorm forecasting system at the Kennedy Space Center
- [Huntrieser, 1995] Heidi I.C. Huntrieser
  Zur Bildung, Verteilung und Vorhersage von Gewittern in der Schweiz
- [Freund, 1997] Y. Freund and R. E. Shapire
  A decision-theoretic generalization of on-line learning and an application to boosting.
- [Ducrocq, 1998] Véronique Ducrocq and Diane Tzanos and Stéphane Sénési
  Diagnostic tools using a mesoscale NWP model for the early warning of convection

[Calas, 2000] C. Calas and V. Ducrocq and S. Sénési
Mesoscale analyses and diagnostic parameters for deep convection nowcasting

[Haklander, 2003] A. J. Haklander and A. Van Delden
Thunderstorm predictors and their forecast skill for the Netherlands

[Burrows, 2005] W. R. Burrows and C. Price and L. J. Wilson
Warm Season Lightning Probability Prediction for Canada and the Northern United States