

Vincent Mirelli and Stephen Tenney

U.S. Army Research Laboratory
2800 Powder Mill Road
Adelphi, MD, 20873-1197

Yoshua Bengio

University of Montreal
P.O. Box 6128, succ. Centre-Ville
Montreal, QC, H3C 3J7, Canada

Nicolas Chapados and Olivier Delalleau

ApSTAT Technologies
4200 Boul. St-Laurent, suite 408
Montreal, QC, H2W 2R2, Canada

Abstract

Machine learning classification algorithms are relevant to a large number of Army classification problems, including the determination of a weapon class from an acoustic signature of a transient. However, much such work has been focused on classification of events from small weapons used for asymmetric warfare, which have been of importance in recent years. In this work we consider classification of very different weapon classes, such as mortar, rockets and RPGs, which are difficult to reliably classify with standard techniques since they tend to have similar acoustic signatures. To address this problem, we compare two recently-introduced state-of-the-art machine learning algorithms, Support Vector Machines and Discriminative Restricted Boltzmann Machines, and develop how to use them to solve this difficult acoustic classification task. We obtain classification accuracy results that could make these techniques suitable for fielding on autonomous devices. Discriminative Restricted Boltzmann Machines appear to yield slightly better accuracy than Support Vector Machines, and are less sensitive to the choice of signal preprocessing and model hyperparameters. Importantly, we also address methodological issues that one faces in order to rigorously compare several classifiers on limited data collected from field trials; these questions are of significance to any application of machine learning methods to Army problems.

The Task

- Use machine learning and advanced signal processing algorithms to discriminate launch signals from three weapon classes: MORTAR, ROCKET, and rocket-propelled grenades (RPGs)
- Many factors affect signal propagation: (1) the distance between receiver and source, (2) the presence of obstacles on the terrain and nature of the ground, (3) the amplitude of the source, (4) the time of day, and (5) the meteorological conditions (cloud cover, wind, and humidity)
- Comparison of two classifiers: Support Vector Machine (SVM: a classical non-parametric discriminant classifier), and Discriminative Restricted Boltzmann Machine (DRBM: a recently proposed hybrid that combines a discriminant criterion and a generative criterion)
- Need a carefully-designed experimental setting in order to properly evaluate the generalization performance of classifiers as if they were deployed on the field

Summary of Available Data

Proving Ground	MORTAR	RPG	ROCKET	Total
APG	197	28	31	256
Dahlgren	0	7	0	7
Yuma	373	0	0	373
Total	570	35	31	636

- Class imbalance (90% MORTAR)
- Low variability in recording conditions (e.g. all ROCKETS are from APG)

Methodological Issues

- Data splits** should mimic the kinds of variations expected between training data and test field data. Two settings considered:
 - REALISTIC-BY-DAY: Test and train recordings cannot come from the same date
 - REALISTIC-BY-RANGE: Test and train recordings cannot come from the same sensor array
 Number of groups of recordings defined by these constraints (randomly assigned to the training or test data by repeated 5-fold cross-validation):

Split Method	MORTAR	RPG	ROCKET
REALISTIC-BY-DAY	5	7	2
REALISTIC-BY-RANGE	11	5	3

- Hyperparameters** cannot be reliably estimated through double cross-validation (low amount of data and splitting constraints) \Rightarrow
 - ANalysis Of VAriance (ANOVA) is performed to determine statistically-significant main effects and interactions
 - Hyperparameters values leading to a performance not statistically significantly different from the best performance are kept
 - Models are compared by the distribution of their performance over the hyperparameters being kept

Classification Algorithm 1: SVMs

Find function f in reproducing kernel Hilbert space \mathcal{H}_K associated to kernel K by:

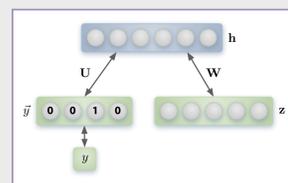
$$\arg\min_{f \in \mathcal{H}_K} \underbrace{C \sum_{z^{(i)} \in \mathcal{I}} (1 - y^{(i)} f(z^{(i)}))_+}_{\text{hinge loss (bias)}} + \frac{1}{2} \underbrace{\|f\|_K^2}_{\text{margin (variance)}}$$

Kernels being used:

- Linear: $K(z^{(1)}, z^{(2)}) = z^{(1)T} z^{(2)}$
- Polynomial: $K(z^{(1)}, z^{(2)}) = (r + z^{(1)T} z^{(2)})^p$
- Radial Basis Function (RBF): $K(z^{(1)}, z^{(2)}) = e^{-\gamma \|z^{(1)} - z^{(2)}\|^2}$

Hyperparameters specific to each kernel are automatically chosen based on an internal three-fold cross-validation on the training set.

Classification Algorithm 2: DRBMs



Restricted Boltzmann Machine modeling the joint distribution of inputs z and target class y

A Discriminative Restricted Boltzmann Machine combines generative and discriminative training criteria by minimizing

$$-\sum_{z^{(i)} \in \mathcal{I}} \log p(y^{(i)}, z^{(i)}) - \lambda \sum_{z^{(i)} \in \mathcal{I}} \log p(y^{(i)} | z^{(i)})$$

where the model's joint probability is

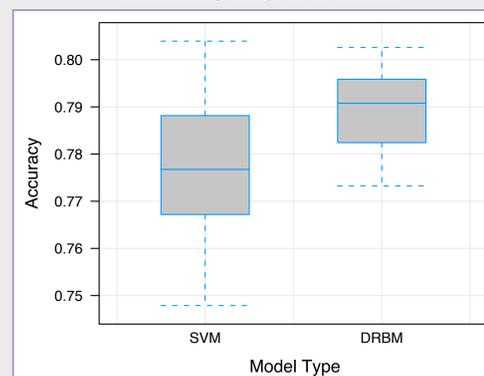
$$p(y, z) \propto \exp(\mathbf{h}^T \mathbf{W} \mathbf{z} + \mathbf{b}^T \mathbf{y} + \mathbf{c}^T \mathbf{h} + \mathbf{d}^T \mathbf{y} + \mathbf{h}^T \mathbf{U} \mathbf{y})$$

Training performed by *contrastive divergence* for the generative part and *stochastic gradient descent* for the discriminative part.

Experimental Setting

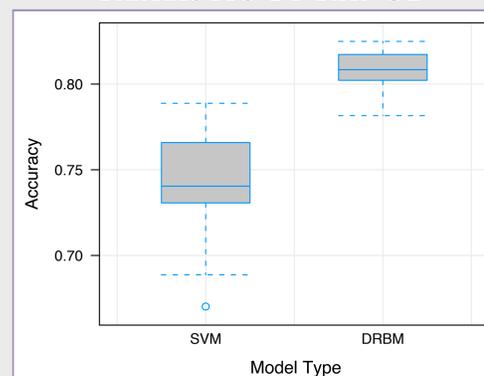
- 5 repetitions of 5-fold cross-validation w.r.t. split constraints defined by the REALISTIC-BY-DAY and REALISTIC-BY-RANGE settings
- Computation of the *normalized* classification accuracy to compensate for class imbalance
- Trying a wide range of hyperparameter values, reporting results for those leading to statistically indistinguishable performance compared to the best

Experimental Results: REALISTIC-BY-DAY



Box-plot over statistically-indistinguishable hyperparameter values for each model type. DRBMs are superior to SVMs both in mean performance (statistically significant) and in robustness (lower variance).

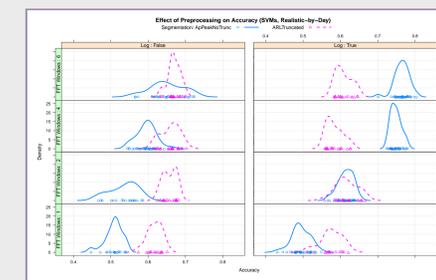
Experimental Results: REALISTIC-BY-RANGE



Box-plot over statistically-indistinguishable hyperparameter values for each model type. DRBMs are superior to SVMs both in mean performance (statistically significant) and in robustness (lower variance).

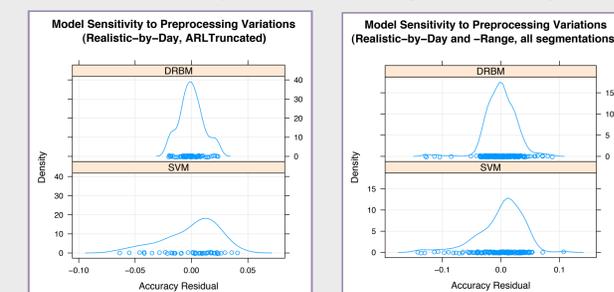
Sensitivity to Preprocessing Choice

- Performance can vary a lot depending on various parameters governing data preprocessing, e.g. type of segmentation, number and size of signal windows for Fast Fourier Transform, etc.



Example of SVM accuracy distribution in the REALISTIC-BY-DAY setting when varying some preprocessing parameters.

- The *accuracy residual* measures the amount of variation in accuracy performance that is due to varying the preprocessing parameters (type of segmentation, number and size of windows for Fast Fourier Transform, ...) while keeping the model hyper-parameters fixed
- We plot the distribution of residuals for two sets of experiments: **on the left** in the REALISTIC-BY-DAY setting and for a single type of segmentation (called ARLTruncated), and **on the right** in the REALISTIC-BY-RANGE setting and for all kinds of segmentations being tried.



- Residuals of SVMs exhibit a greater variance than those of DRBMs (difference is statistically significant)
- Lower variability \Rightarrow performance is more reliably estimated \Rightarrow DRBMs are particularly useful when not much data is available

Conclusion

- Tackling a challenging detonation type classification task
- Proposed methodology to properly train and evaluate classifiers
- Application to SVMs and DRBMs
- DRBMs slightly superior overall, and less sensitive to the choice of preprocessing hyperparameters than SVMs.

- Bengio, Y. (2009). Learning Deep Architectures for AI. *Foundations and Trends in Machine Learning*, to appear.
- Cortes, C. and V. Vapnik (1995). Support-Vector Networks. *Mach. Learn.* 20(3), 273–297.
- Larochelle, H. and Y. Bengio (2008). Classification using discriminative restricted Boltzmann machines. In A. MCCALLUM and S. ROWEIS (Eds.), *Proceedings of the 25th Annual International Conference on Machine Learning (ICML 2008)*, pp. 536–543. Omnipress.