Science & Technology 2013, Poster T3-P67

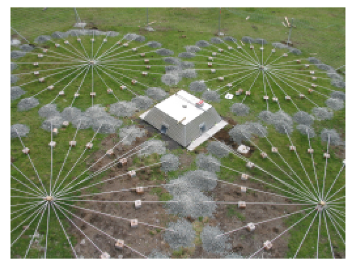
Matthias Tuma[1,2] Christian Igel[2]
1) Institut für Neuroinformatik (INI), Ruhr-Universität Bochum 2) Department of Computer Science (DIKU), University of Copenhagen

Investigation of Kernel-Based Machine Learning Techniques for Infrasound Signal Classification

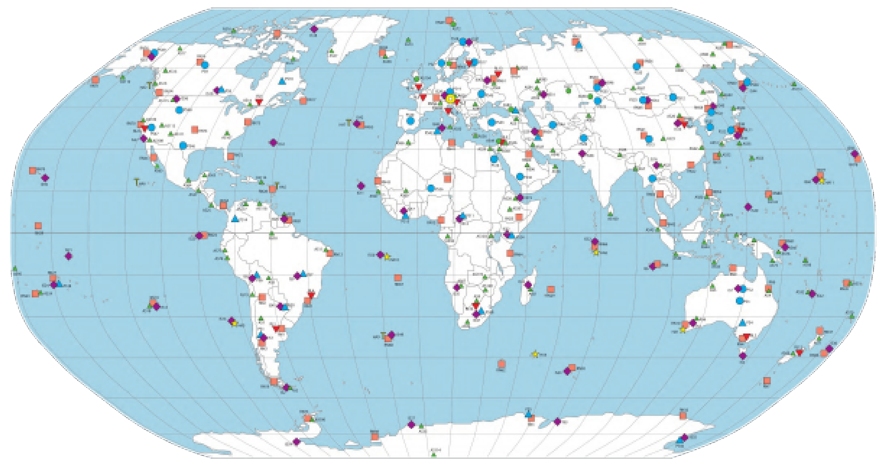
RUB

1. Background

For future verification of the Comprehensive Nuclear-Test-Ban Treaty, a global sensor network continuously records infrasound signals at stations around the world.



Infrasound station IS49, Tristan da Cunha, United Kingdom Source: CTBTO



Map of IMS stations (including auxiliary stations). Infrasound stations are marked by a purple square. Source: CTBTO

Infrasound stations pick up lower-frequency pressure variations in the atmosphere. IMS infrasound stations mainly target detection of atmospheric nuclear explosions.

The present project plans to explore kernel-based machine learning algorithms (in particular, support vector machine classifiers) for two related discrimination tasks in station-level infrasound processing.

2. Infrasound Event Types

Source 1,2,4,5: Wikimedia
Source 3: CTBTO

Infrasound producing events include (marine) storms, earthquakes, meteors, and volcanoes; as well as civil and military blasts and explosions, aircraft, and rocket launches as man-made causes. Because of a propensity of small, irrelevant, repetitive, local and/or station-specific background noises to clutter IDC bulletins with false associations/events, IDC processing includes a detection categorization phase.

3. Infrasound Propagation

Travel path and mode for infrasound vary greatly with temperature, wind pattern, as well as daily and annually by variations in atmospheric conditions. Infrasound signals can be broadly discriminated according to their travel path in the atmosphere: depending on the maximum height of reflection, one can distinguish between tropospheric, stratospheric, and thermospheric arrivals.

Atmospheric layers as seen from the ISS.
Source: NASA EOL

Disclaimer

The views expressed herein are those of the authors and do not necessarily reflect those of the CTBTO Preparatory Commission.

Contact

matthias.tuma@rub.de
Institut für Neuroinformatik
Ruhr-Universität Bochum, Germany

4. IDC Infrasound Processing

Three stages can be identified in IDC infrasound processing: automated processing on the single-array level; automated processing on the network level; and interactive analyst review. On the station level, detection and feature extraction is done via PMCC (Cansi 1995). After detection and feature extraction, a detection categorization stage attempts to filter out noises against relevant signals/phases. Here, a noise type detection is anything which would not be possible to associate to any other IMS waveform station's detection. Predictions within this categorization stage guard automatic and analyst processing against an overwhelming amount of spurious detections and associations.

An (optional) additional categorization step concerns the travel path of the infrasound arrival. Phases can be discriminated into tropospheric, stratospheric, and thermospheric arrivals.

5. Project Overview

The present project will investigate kernel-based machine learning algorithms, in particular support vector machines (SVMs, Cortes & Vapnik 1995), for the automatic classification of infrasound signals. This includes categorization of incoming detections into two classes (signal or noise) and identification of travel path (tropospheric, stratospheric, or thermospheric). The project will thus develop a processing module situated after PMCC processing and before global association and event building. The methodology can in parts follow that of a sibling project for hydroacoustic signal classification (Tuma et al. 2012)

6. Data Basis

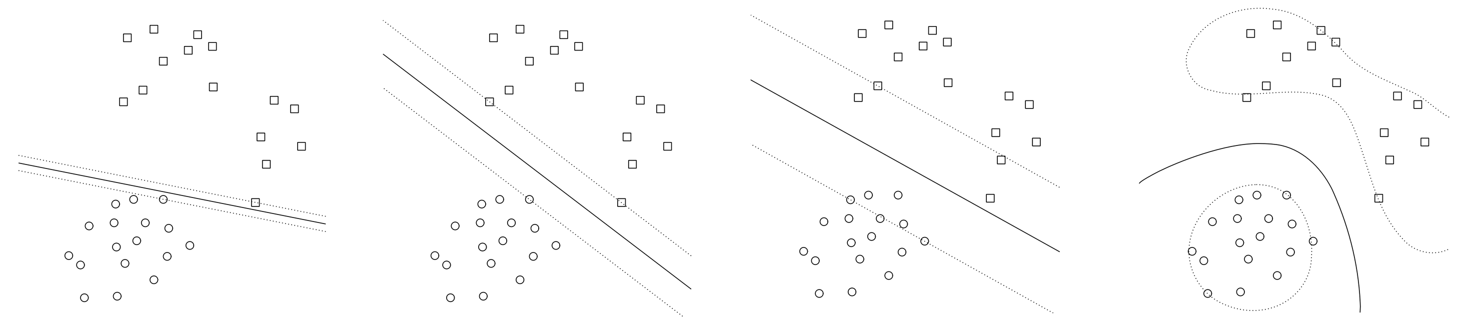
At this point, there is no pre-compiled reference dataset available to train, evaluate and compare classification algorithms for infrasound phase type and travel path discrimination. Notable sources for reference infrasound events can be (1) the Infrasound Reference Database (IRED); and (2) the virtual Data Exploitation Centre (vDEC). The IRED contains over 750 reference entries from mine and quarry blasts, other explosions, earthquakes, volcano activity, rocket launches, meteorites, aircraft, and other sources. The vDEC provides general access to IDC raw waveforms and derived products. Extracting a custom set of reference signals from vDEC or mixing IRED and signals extracted from vDEC are alternative options.

7. Kernel-based Classification

In the supervised learning scenario, adaptation (or: construction) of a classifier is automatically driven by sample data S (cf. previous section). Let $S = \{(x_i, y_i) \,|\, 1 \leq i \leq \ell\}$ be drawn from an unknown distribution p over $X \times Y$, with X the input and Y the output set. Abstractly speaking, the goal of binary classification would be to infer a hypothesis $h : X \to Y$ that minimizes the expected risk

$$R_p(h) = \int_{X \times Y} L_{0-1}(y, h(x))\, dp(x, y) \ ,$$

where $L_{0-1}(y, z)$ is 0 if $y = z$ and 1 otherwise (i.e., the 0-1-loss). Simply put, the classifier should make as few mistakes as possible considering the overall distribution.



(a) Linear separation by a small margin. (b) The linear large-margin solution. (c) Also allowing for margin violations. (d) Non-linear separation with outliers.
Figure 1: A linearly separable two-class dataset, and different hypotheses for a discrimination boundary between them, as can be found by SVM classifiers.

As p is unknown, the expected risk is in practice estimated using the sample data S. To avoid overfitting (i.e., learning idiosyncrasies of S instead of properties of p), a regularization term is added that penalizes complex solutions. Also, the 0-1 loss is difficult to optimize and thus replaced by a convex relaxation.

Support vector machines construct a classification hypothesis as a linear expansion of the training data in a kernel-induced feature space. That is, consider a positive semi-definite (Mercer) kernel function $k : X \times X \to \mathbb{R}$. Then, SVMs build on the feature space $\mathcal{H}_k = \text{span}\{k(x, \cdot) \,|\, x \in X\}$ and the function class $\mathcal{H}_k^b = \{f = g + b \,|\, g \in \mathcal{H}_k, b \in \mathbb{R}\}$. The decision boundary induced by the sign of a function $f \in \mathcal{H}_k^b$ is a hyperplane in $\mathcal{H}_k$. 1-Norm Soft Margin SVMs find such a hyperplane as a solution to a quadratic optimization problem which results from the paradigm of regularized risk minimization introduced above – that is, combining the aim for low training error with one for a simple hypothesis:

$$\underset{f \in \mathcal{H}_k^b}{\text{minimize}} \quad \frac{1}{\ell} \sum_{i=1}^{\ell} L_{\text{hinge}}(y_i, f(x_i)) + (2\ell C)^{-1} \|f\|_k^2 \ ,$$

with loss function $L_{\text{hinge}}(y, f(x)) = \max\{0, 1 - yf(x)\}$. The parameter $C > 0$ controls the trade-off between reducing the empirical loss L_{hinge} and the complexity of the hypothesis, as measured by its norm $\|.\|_k$.

8. Model Selection and Processing Optimization

To optimize SVM hyperparameters for binary classification problems (here: signal-noise phase ID task) using standard kernels, baseline approaches like grid search on the cross-validation error can be employed. In addition, gradient-based methods like maximum-likelihood model selection are available for more flexible kernels (e.g., Glasmachers and Igel 2010). Following (Tuma 2012), parameters of pre-processing may be included in the overall model optimization.

For multi-class problems (here: path ID task and/or joint phase/path ID task), a number of multi-class SVM formulations are available along with fast solver implementations (Dogan et al. 2011). All methods are available and experiments will be implemented in the Shark machine learning library (Igel et al. 2008).

References

[1] Brachet, N; Brown, D; Mialle, P; Le Bras, R: Infrasound data processing for CTBT verification – sources, CTBTO ISS Conference, Vienna, 10-12 June 2009.

[2] Brachet, N; Brown, D; Mialle, P; Le Bras, R: Infrasound data processing for CTBT verification – station processing, CTBTO ISS Conference, Vienna, 10-12 June 2009.

[3] Cansi Y: An Automatic Seismic Event Processing for Detection and Location: The P.M.C.C. Method, Geophys. Res. Lett., 22, pp. 1021–1024, 1995.

[4] Cortes, C; Vapnik, V: Support-Vector Networks. Machine Learning, 20, pp. 273–297, 1995.

[5] Tuma, M; Igel, C; Prior, M: Hydroacoustic Signal Classification Using Support Vector Machines. In: Chen, C. (ed.): Signal and Image Processing for Remote Sensing, 2nd edition, CRC Press, pp. 37–65, 2012.

[6] Glasmachers, T; Igel, C: Maximum Likelihood Model Selection for 1-Norm Soft Margin SVMs with Multiple Parameters. IEEE Transactions on Pattern Recognition and Machine Intelligence 32: 1522–1528, 2010.

[7] Dogan, Ü; Glasmachers, T; Igel, C: Fast Training of Multi-Class Support Vector Machines. Technical Report, Department of Computer Science, University of Copenhagen, 2011/3, 2011.

[8] Igel, C; Heidrich-Meisner, V; Glasmachers, T: Shark. Journal of Machine Learning Research, 9:993–996, 2008.