Geophysical Research Abstracts
Vol. 16, EGU2014-7720, 2014
EGU General Assembly 2014

Kernel-based machine learning techniques for infrasound signal classification

Matthias Tuma (1,2), Christian Igel (2), and Pierrick Mialle (3)

(1) Institut für Neuroinformatik, Ruhr-Universität Bochum, Germany (matthias.tuma@rub.de), (2) Department of Computer Science, University of Copenhagen, Denmark (igel@diku.dk), (3) CTBTO Preparatory Commission, Vienna, Austria (pierrick.mialle@ctbto.org)

Infrasound monitoring is one of four remote sensing technologies continuously employed by the CTBTO Preparatory Commission. The CTBTO's infrasound network is designed to monitor the Earth for potential evidence of atmospheric or shallow underground nuclear explosions. Upon completion, it will comprise 60 infrasound array stations distributed around the globe, of which 47 were certified in January 2014.

Three stages can be identified in CTBTO infrasound data processing: automated processing at the level of single array stations, automated processing at the level of the overall global network, and interactive review by human analysts. At station level, the cross correlation-based PMCC algorithm is used for initial detection of coherent wavefronts. It produces estimates for trace velocity and azimuth of incoming wavefronts, as well as other descriptive features characterizing a signal. Detected arrivals are then categorized into potentially treaty-relevant versus noise-type signals by a rule-based expert system. This corresponds to a binary classification task at the level of station processing. In addition, incoming signals may be grouped according to their travel path in the atmosphere.

The present work investigates automatic classification of infrasound arrivals by kernel-based pattern recognition methods. It aims to explore the potential of state-of-the-art machine learning methods vis-a-vis the current rule-based and task-tailored expert system. To this purpose, we first address the compilation of a representative, labeled reference benchmark dataset as a prerequisite for both classifier training and evaluation. Data representation is based on features extracted by the CTBTO's PMCC algorithm. As classifiers, we employ support vector machines (SVMs) in a supervised learning setting. Different SVM kernel functions are used and adapted through different hyperparameter optimization routines. The resulting performance is compared to several baseline classifiers. All experiments are conducted using the Shark machine learning library.