Computers & Geosciences 87 (2016) 22–30

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

Automatic classification of seismic events within a regional seismograph network

Jari Kortström*, Marja Uski, Timo Tiira

Institute of Seismology, University of Helsinki, POB 68, FI-00014, Finland

ARTICLE INFO

Article history:
Received 18 May 2015
Received in revised form
14 September 2015
Accepted 9 November 2015
Available online 14 November 2015

Keywords:
Classification
Regional seismic events
Automatic method
Support Vector Machine
Sparse seismic network
Fennoscandia

ABSTRACT

This paper presents a fully automatic method for seismic event classification within a sparse regional seismograph network. The method is based on a supervised pattern recognition technique called the Support Vector Machine (SVM). The classification relies on differences in signal energy distribution between natural and artificial seismic sources. We filtered seismic records via 20 narrow band-pass filters and divided them into four phase windows: P, P coda, S, and S coda. We then computed a short-term average (STA) value for each filter channel and phase window. The 80 discrimination parameters served as a training model for the SVM. We calculated station specific SVM models for 19 on-line seismic stations in Finland. The training data set included 918 positive (earthquake) and 3469 negative (non-earthquake) examples. An independent test period determined method and rules for integrating station-specific classification results into network results. Finally, we applied the network classification rules to independent evaluation data comprising 5435 fully automatic event determinations, 5404 of which had been manually identified as explosions or noise, and 31 as earthquakes. The SVM method correctly identified 94% of the non-earthquakes and all but one of the earthquakes.

The result implies that the SVM tool can identify and filter out blasts and spurious events from fully automatic event solutions with a high level of accuracy. The tool helps to reduce the work-load and costs of manual seismic analysis by leaving only a small fraction of automatic event determinations, the probable earthquakes, for more detailed seismological analysis. The self-learning approach presented here is flexible and easily adjustable to the requirements of a denser or wider high-frequency network.

1. Introduction

Many seismological observatories use automatic event detection and location procedures for monitoring local and regional seismicity. Fully automatic event solutions provide a cost-effective, nearly real-time snapshot on seismic activity within a target area. However, the events are often unclassified or poorly classified. The next and crucial step is to apply reliable automatic or semi-automatic methods for classifying the huge database of fully automatic event solutions. Automated event classification is necessary in monitoring natural hazards when rapid and reliable information to local authorities and media is of essence. Moreover, it helps to maintain the quality of regional earthquake catalogs, in particular among the low magnitude events. Namely, if unclassified or poorly classified event solutions end up in the catalog, the earthquake data will become increasingly contaminated with anthropogenic activity. Investigations relying upon such data will yield erroneous estimates of the rate of seismicity and, consequently, of seismic hazard.

The classification of seismic events requires the integration of physical and statistical techniques. The task is challenging in low-seismicity areas where natural and anthropogenic seismicity often overlap in magnitude, space and time. A sparse coverage of the monitoring network further complicates event classification. The Finnish National Seismic Network, operated by the Institute of Seismology, University of Helsinki (ISUH) is a typical example of a sparse regional network. To supplement the near real time automatic detection and location capability of the national network, ISUH utilizes also available on-line stations of the partner networks (Fig. 1). The area monitored by ISUH covers central and eastern parts of Fennoscandia, including Finland, parts of Estonia, Norway, Sweden, Russia and the adjoining seas (Fig. 1). The region is characterized by a relatively low rate of natural seismicity intermingled with a high rate of anthropogenic activity. Several large-scale underground mines along with numerous open pits operate in the area on a daily basis. The greatest mine blasts have magnitudes exceeding M_L 3. Rock bursts and other mining-

* Corresponding author.
E-mail address: Jari.Kortstrom@helsinki.fi (J. Kortström).

http://dx.doi.org/10.1016/j.cageo.2015.11.006
0098-3004/

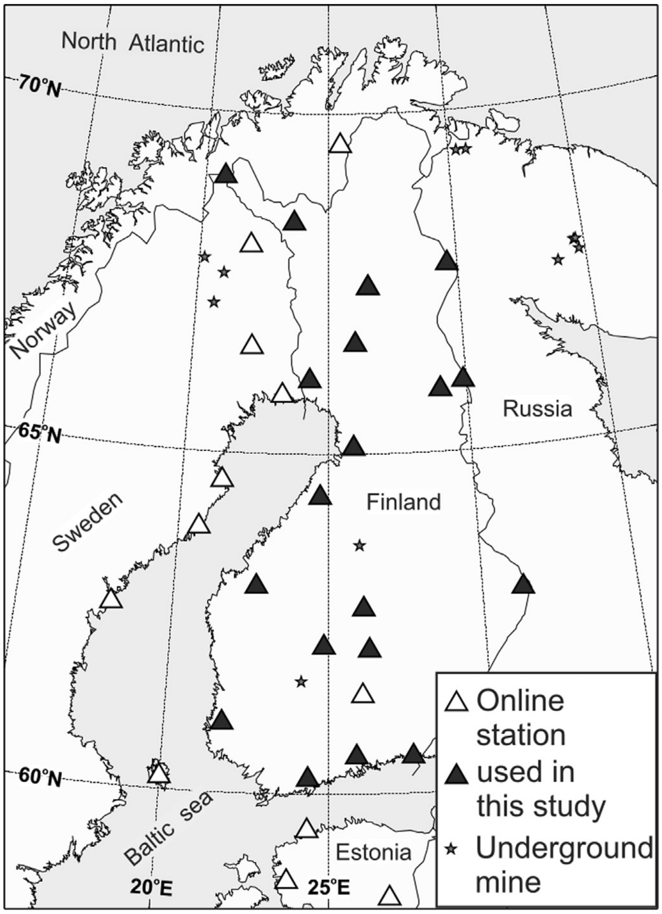

Fig. 1. A map showing online seismograph stations used for automatic detection and location of regional seismic events. High-frequency stations used for event classification are filled in black. The locations of underground mines are included for comparison.

induced or -triggered events occur frequently, the largest recorded so far are of magnitude M_L 4+ (e.g., Roth and Bungum (2003)). Continual explosions from e.g. military exercises present a challenge to seismic monitoring of the sea areas.

In addition to fully automatic near real-time bulletins, ISUH releases reviewed event bulletins where the automatic event solutions have been manually reviewed: all events are classified, spurious events are cleaned out and possible earthquakes (~1% of the data) as well as events with questionable seismic origin are subjected to a detailed reanalysis. The visual screening phase is time-consuming and labor-intensive and calls for automation.

Many automatic seismogram classification methods reduce the waveform data to a set of parameters and these parameter vectors are then classified. Parameters commonly used in classification of regional events are spectral amplitude ratios of different seismic phases, complexity of the signal and autoregressive moving average (ARMA) coefficients. Recent examples are e.g. Fäh and Koch (2002), Zeiler and Velasco (2009) and Yilmaz et al. (2013). More complex methods have also been applied. Allmann et al. (2008) compared spectral fit to a simple ω^{-2} source model, whereas Lyubushin et al. (2013) used multi-fractal singularity spectrum properties.

Our study originates from an idea to translate the guidelines for manual spectral analysis into automatic classification parameters. According to our experience the time-frequency distribution plots, i.e., spectrograms, are the most powerful tool in discriminating weak local and regional events. We exploit the total duration of seismic signals by forming "numerical spectrograms" of the automatically detected and located events. In order to present the information contained in the spectrogram plots in numerical form, a large parameter set is needed. We therefore search for a classification method that is effective in high dimensional spaces.

Both statistical and machine learning methods have been applied in seismic classification previously. Examples of statistical methods are linear Bayesian discriminator (Lyubushin et al., 2013), linear discrimination analysis and its variants (Che et al., 2007; Kuyuk et al., 2014) and multivariate statistical analysis (Fäh and Koch, 2002). Examples of machine learning methods include the use of supervised Artificial Neural Networks (ANN) (Tiira, 1996; AllamehZadeh, 2011). Kuyuk et al. (2011) have used Self Organizing Maps and ANN combined with unsupervised learning in classification of small earthquakes and quarry blasts. Two unsupervised machine learning methods, k-means and Gaussian mixture model, were applied by Kuyuk et al. (2012) for classification of seismic activities in Istanbul. Support Vector Machine (SVM) is a popular application for a wide range of supervised pattern recognition problems (e.g., Boser et al., 1992; Cortes and Vapnik, 1995; Vapnik, 1995). Giacco et al. (2009) have applied SVM to automatic classification of seismic signals in volcano environment and Zhao (2007) for seismic discrimination within hydrocarbon reservoirs. The advantages of using SVM are handleability of large number of features and effectiveness in high dimensional spaces. SVM also gives unambiguous result to an ambiguous problem, which is easily implementable into automatic processing.

Our approach, the numerical spectrogram, is a set of parameters calculated over time and frequency space of seismic records. Classification of the parameter set, in turn, is basically a pattern recognition problem (Joachims, 1999). For solving the problem, we have chosen to use the SVM^{light} package (svmlight.joachims.org), which is an implementation of Vapnik's Support Vector Machine (Vapnik, 1995).

Section 2 summarizes the basis for our parameterization, i.e. the guidelines used in visual spectral analysis. In Section 3 we present data and methods applied to design an automatic SVM tool capable of identifying and filtering human-made and spurious events from automatic seismic event bulletins with a high level of accuracy. The goal is that only a small fraction of the events, i.e., the probable earthquakes, are left for manual screening and revision. We will apply the tool to fully automatic regional seismic event solutions produced by ISUH and we will show that the SVM based tool performs well within the network setting and relevant boundaries.

2. Spectral features of typical earthquakes and explosions

Manual discrimination of seismic events relies on judgments made by individual analysts. To increase objectivity in the decision making ISUH has listed the following guidelines for visual seismogram analysis.

Earthquakes are volume sources extended both in time and space and they generate a larger fraction of energy in S waves than in P waves. The P and S wave radiation patterns are, however, strongly dependent on rupture directivity. For earthquake sources the strength of P and S wave signals may vary significantly at stations located at approximately the same epicentral distance but in different azimuth directions. Seismic waves of earthquakes have wide frequency content and their energy is evenly distributed over the whole recorded frequency band. Earthquakes also produce rather complex waveforms because of secondary depth-sensitive seismic phases in their P and S coda.

In contrast to earthquakes, explosions are compressive point sources from which P wave energy radiates evenly to all azimuth directions. S waves are presumably generated by mode