AUTOMATICALLY FINDING AVALANCHES IN GEOPHONE DATA: A PATTERN RECOGNITION WORKFLOW

Marc J. Rubin*[1], Tracy Camp[1], and Alec van Herwijnen[2]

[1]Dept. of Electrical Engineering and Computer Science, Colorado School of Mines, Golden, CO, USA
[2]WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland

ABSTRACT: In this article we summarize a pattern recognition workflow to automatically detect avalanche events from passive seismic data collected from geophones near Davos, Switzerland during the 2010-2011 snow season. Our workflow consists of three steps: 1) spectral flux based event selection, 2) feature extraction, and 3) classification. The results are quite promising: our workflow achieves 93% overall classification accuracy with 13% precision for detecting avalanches for the entire season.

1. INTRODUCTION

Automatically detecting avalanches in near real-time (and in any visibility) would provide avalanche forecasters and highway crews with very important information to help make informed decisions regarding avalanche danger or road closures. In this article we describe a pattern recognition workflow to automatically detect avalanches from passive seismic (geophone) data. In particular, we describe the signal processing and machine learning techniques we used to detect avalanches from geophone data collected during the 2010-2011 winter season near Davos, Switzerland.

Previous researchers have used pattern recognition algorithms to detect avalanches from preprocessed seismic data. Most notably, the SARA (System for Avalanche Recognition Analysis) software suite uses manually trained fuzzy logic rules to identify avalanches (Leprettre et al. (1996, 1998a,b); Navarre et al. (2009)). The downside to their approach is that the fuzzy logic rules are based on manual (expert) analysis of previous data; in other words, the rules are not derived automatically or in a timely manner.

Bessason et al. (2007) used a distanced weighted k-nearest neighbor approach (k=3) on previously recorded seismic data. The classification results of this automated method were unsatisfactory; specifically, only 65% (78 of 119) of the avalanches were correctly identified using their k-nearest neighbor algorithm approach. This work suggests that there is considerable room for improvement for automated avalanche detection.

* *Corresponding author address:* Marc J. Rubin, Dept. of Electrical Engineering and Computer Science, Colorado School of Mines, Golden, CO, USA 80401; tel: (713) 628 - 5822; email: mrubin@mines.edu

2. PATTERN RECOGNITION

In this section we describe the pattern recognition workflow we used. First, we briefly describe the seismic data set collected from geophones. Second, we detail how we used spectral flux based event selection to pick sizable events of interest. Third, we highlight the 10 features we extracted from the frequency domain and subsequently use for event classification. Lastly, we summarize our experimentation with 12 different classification algorithms trained and tested on the seismic data.

2.1. Geophone Data

The data set consisted of seismic data collected during the 2010-2011 snow season from seven geophones located in a snow slope near Davos, Switzerland. The geophones recorded data at 500 Hz with 24-bit precision for over 100 days. More details regarding the deployment can be found in Herwijnen and Schweizer (2011).

Within the seismic data, 385 possible avalanches were identified, ranging from three seconds to nearly two minutes in length. Of the 385 possible avalanches, 33 were considered large avalanches while the remaining 352 were assumed to be small avalanches (i.e., sluffs). Our pattern recognition workflow focused mainly on positively identifying the 33 large events, i.e., we felt it was acceptable to miss some of the smaller events in favor of improved results for detecting the larger events.

The seismic data was far from clean. There was much background and spurious noise caused by a variety of sources: e.g., wind, ski lifts, snow cats, avalanche bombing, helicopters, airplanes, earthquakes, etc. The next section discusses how we processed the noisy data.

2.2. Spectral Flux Based Event Selection

To select only events of interest from the seismic data set, we used spectral flux to determine five second frames with significant instantaneous increases in spectral energy. Spectral flux is simply the Euclidean distance between all points in two consecutive spectral frames (a 2048-bin, non-overlapping fast Fourier transform (FFT)). An event was selected if the instantaneous energy was above a predetermined percentage threshold.

In our workflow, we chose a threshold of 90%, meaning that a five-second frame was selected if the spectral flux increased by 90% relative to the surrounding five minutes of data (Figures 1 and 2). Using this method, we selected 32 of 33 slabs, 246 of 352 sluffs, and 32,544 non-avalanche events.

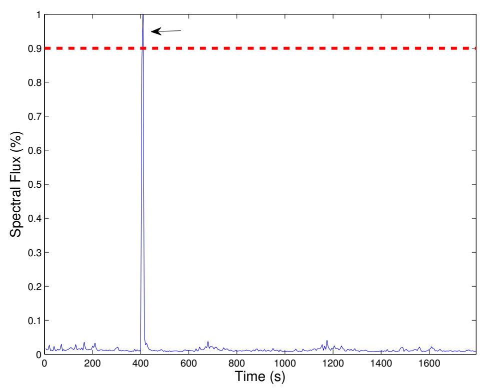

Figure 1: The spectral flux of a slab avalanche event that occurred on January 22, 2011 (at time 400 seconds). The red line represents the 90% threshold.

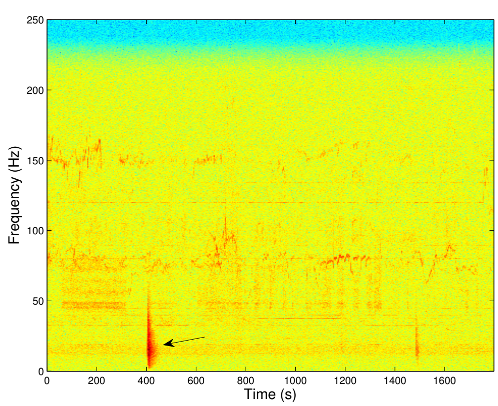

Figure 2: The frequency domain of a slab avalanche that occurred on January 22, 2011 (at time 400 seconds).

2.3. Feature Extraction

The next step was to transform each five-second selected event into quantifiable features that differentiate the avalanches from the non-avalanche events. Using an open-source Matlab toolbox created for music signal processing (i.e., MIRToolBox developed by Lartillot and Toiviainen (2007)), we extracted 10 features from the frequency domain (Table 1). These features, often used in music pattern recognition (e.g., Klapuri and Davy (2006)), provide a numerical summary of the size, shape, and peak of the frequency spectrum (e.g., Figure 3). It is important to note that the relatively slow sampling rate (500 Hz) and close proximity of the seven geophones (five to 10m) made estimating characteristics of the seismic waveform (e.g., velocity, arrival times, back azimuth, etc.,) implausible.

Source	**Features**
Top 1% Energy	mean, standard deviation, maximum
Frequency Domain	centroid, spread, skewness, regularity, flatness, 85% rolloff, kurtosis

Table 1: We extracted 10 features from the frequency domain to create a quantitative summary of each five-second spectral frame. Three features (i.e., mean, standard deviation, and maximum) were calculated from the 1% most powerful frequencies of each frame.

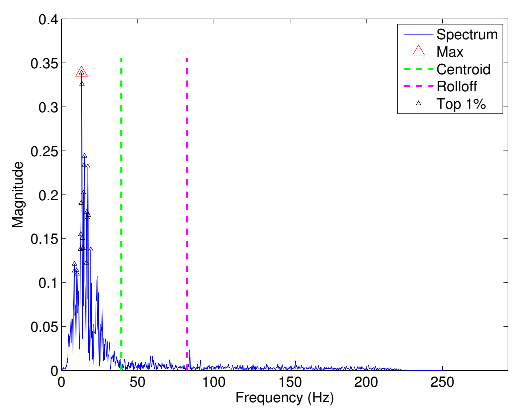

Figure 3: For each five second frame, we calculated several features from the frequency spectrum (2048-bin FFT). This event is a slab avalanche recorded on January 22, 2011.

2.4. Classification

The last and most important step in our pattern recognition workflow was to build a model to detect the 278 avalanche frames from the 32,822 total selected events. To do this, we experimented with 12 different classification algorithms ranging from probabilistic and statistical (e.g., Bayes, Gaussian processes), to highly non-linear function approximators (e.g., artificial neural network, support vector machine). Specifically, we tested an artificial neural network (ANN), naive Bayes, Bayes network, CART tree, fuzzy logic rules, Gaussian processes (Gauss), J48 Tree, k-nearest neighbors (KNN), random forest (RanForest), RIPPER, decision stump, and support vector machine (SVM).

The classification experiments consisted of 100 iterations of training and testing, where training was performed using 10% of all known avalanches (both slabs and sluffs) and an equal number of non-avalanche events. In each iteration, all remaining data not used for training was used to test the classifier.

The non-avalanche events were selected using stratified cluster-based subsampling, which is a method used to mitigate the unfavorable effects of extreme class imbalance issues (Yen and Lee (2009)). Briefly, we used K-means clustering to separate the non-avalanche events into seven different groups, and then employed stratified subsampling to insure fair representation of each cluster in the training subset. We chose seven groups because there were roughly seven types of noise events: i.e., wind, ski lifts, snow cats, airplanes, helicopters, avalanche control work (bombing), and earthquakes.

The results from our experiments are quite promising (Table 2). All 12 algorithms had 80% overall classification accuracy or above, with 10 performing over 90%. Furthermore, all classifiers (with the exception of KNN) reported precision rates at or above 7%, with seven achieving precision rates over 10%. The best classifier was a decision stump, which reached 93% overall accuracy, nearly 90% recall, and over 13% precision for the entire season's worth of data. We note that neither Leprettre et al. (1998a) nor Bessason et al. (2007) report the precision of their classification models.

3. CONCLUSIONS

In this article we present our successful pattern recognition workflow to detect avalanches from geophone data. There are several conclusions that can be made from our results. Most notably, using data from only a single geophone sensor, we can detect avalanches with over 90% accuracy and 13% precision. For the avalanche forecasting practitioner, these results imply that a single geophone may be a viable and inexpensive option to monitor specific avalanche paths. Additionally, with the rapid improvements in wireless sensing technology, the possibility of using inexpensive wireless geophones to detect avalanches is becoming a reality.

Algorithm	Accuracy*	Recall†	Precision‡
STUMP	**93.0**	**89.5**	**13.2**
SVM	92.9	89.7	10.5
CART	92.4	89.7	12.3
Gauss	92.4	91.2	09.9
RanForest	92.1	89.5	10.2
RIPPER	91.6	88.8	11.7
Bayes	91.3	90.3	10.4
J48	90.9	88.6	10.7
BayesNet	90.4	89.9	07.9
ANN	90.3	88.1	08.1
Fuzzy	89.7	90.5	08.2
KNN	81.4	82.8	03.6

Table 2: The mean results of 12 machine learning algorithms trained and tested 100 times on a single sensor's data. The results are sorted by accuracy.

There are two future directions we plan to take in our work. First, for the 2012-2013 snow season, we plan to deploy the seven wired geophones in a circular pattern with a radius of 30m (similar to Lacroix et al. (2012)). With such an arrangement, we should be able to estimate the seismic waveforms of each event. In addition, by estimating arrival times, we should be able to calculate an avalanche's seismic velocity and direction of travel; this information will be very helpful in differentiating ambiguous signals.

Second, in addition to the circular arrangement of wired geophones, we also plan to install an inexpensive wireless geophone in the same location. The wireless geophone is a prototype system developed by the Colorado School of Mines to record continuous geophone data at 250 Hz with 24-bit precision and 64X gain (Figure 4). Initial field tests revealed that our inexpensive wireless prototype (˜$100 excluding sensor) performs nearly identically to a $750 single channel digitizer (i.e. Cirrus Logic CRD5378 excluding sensor). Results from the 2012-2013 field deployment will inform the viability of our inexpensive wireless platform for detecting avalanches.

*Accuracy: Percentage of all frames correctly predicted.

†Recall: Percentage of avalanches correctly predicted.

‡Precision: Percentage of frames labeled as avalanches that were actually avalanches.

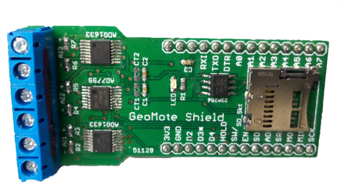

Figure 4: The GeoMoteShield, designed by Colorado School of Mines, will be used to interface a geophone sensor with an Arduino Fio wireless mote platform.

4. REFERENCES

Bessason, B., G. Eiriksson, O. Thorarinsson, A. Thorarinsson, and S. Einarsson, 2007: Automatic detection of avalanches and debris flows by seismic methods. *Journal of Glaciology*, **53 (182)**, 461–472.

Herwijnen, A. and J. Schweizer, 2011: Monitoring avalanche activity using a seismic sensor. *Cold Regions Science and Technology*, **69 (2-3)**, 165–176.

Klapuri, A. and M. Davy, 2006: *Signal Processing Methods for Music Transcription*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Lacroix, P., J. Grasso, J. Roulle, G. Giraud, D. Goetz, S. Morin, and A. Helmstetter, 2012: Monitoring of snow avalanches using a seismic array: Location, speed estimation, and relationships to meteorological variables. *Journal of Geophysical Research*, **117 (F01034)**.

Lartillot, O. and P. Toiviainen, 2007: A Matlab toolbox for musical feature extraction from audio. *Proceedings of the 10th International Conference on Digital Audio Effects*.

Leprettre, B., N. Martin, F. Glangeaud, and J.-P. Navarre, 1998a: Three-component signal recognition using time, time-frequency, and polarization information– application to seismic detection of avalanches. *IEEE Transactions on Signal Processing*, **46 (1)**, 83 –102, doi:10.1109/78.651183.

Leprettre, B., J.-P. Navarre, J. Panel, F. Touvier, A. Taillefer, and J. Roulle, 1998b: Prototype for operational seismic detection of natural avalanches. *Annals of Glaciology*, **26**, 313–318.

Leprettre, B., J.-P. Navarre, and A. Taillefer, 1996: First results from a pre-operational system for automatic detection and recognition of seismic signals associated with avalanches. *Journal of Glaciology*, **42 (141)**, 352–363.

Navarre, J.-P., E. Bourova, J. Roulle, and D. Y. Delio, 2009: The seismic detection of avalanches: an information tool for the avalanche forecaster. *Proceedings of the International Snow Science Workshop*.

Yen, S. and Y. Lee, 2009: Cluster-based undersampling approaches for imbalanced data distributions. *Expert Systems with Applications*, **36**, 5718–5727.