SIAMESE UNSUPERVISED CLUSTERING FOR REMOVING UNCERTAINTY IN MICROSEISMIC SIGNAL LABELLING

David Murray, Lina Stankovic, Vladimir Stankovic[*]

Electronic and Electrical Engineering
University of Strathclyde
Glasgow, UK

ABSTRACT

The labelling of large seismic datasets is a challenging problem. Currently the methods most favoured by geoscientists are based on well known geophysical properties with STA/LTA ratio pickers remaining highly trusted to generate results which can be quickly attributed due to their ability to pick relatively high Signal to Noise Ratio (SNR) events with high speed and accuracy. We aim to improve on the ability of deep learning methods by the unsupervised clustering of events which can help to visually identify results as belonging to a certain cluster with high confidence without the need for event by event processing. From our previous work we use a Siamese model trained with known labels from an open source dataset we show performance as a classifier and then expand on the method by showing clustering of events, where an expert can have high confidence that certain events are correctly identified, or require further evaluation.

Index Terms— Siamese Network, Microseismic, Unsupervised Clustering, Self-ordering Maps

1. INTRODUCTION

The detection of seismic events has generally focused on larger scale events such as earthquakes and volcanic eruptions. More recently with advances in sensing and availability, more sites are being monitored at global and regional scales than ever before. This increase means that much more data needs to be investigated and catalogued to help better understand the underlying geophysical processes. With advances in hydraulic fracturing and changes in climate more attention is being paid to sites where the possibility of induced seismicity, landslides and rockfalls may occur. These type of events can be sudden and unpredictable as they may not be preceded by a large earthquake, as such being able to detect precursors at the microseismic level is becoming increasingly important. While larger seismic events such as earthquakes (M>3) are well understood from a detection point of view with usually easily identifiable P and S waves, micro seismic events are much less likely to be identified when it comes to pickers as they tend to have a low SNR and short duration. Therefore cataloguing these seismic events is a highly time consuming process and unless the event is significant (in term of signal to noise ratio (SNR)) then quickly finding and classifying it becomes non-trivial. Microseismic events can be extremely frequent, and labelling can become prone to error in the presence of background anthropogenic noise from machinery and wildlife.

Machine learning, and specifically deep learning, have become prevalent within the Earth Sciences community. Indeed, the scale of this is highlighted in a recent review paper [1], which breaks down the various applications that machine learning is being used for - microseismic event detection, source localisation [2, 3], and cluster analysis [4] being the most prominent. For recent approaches, using hand-crafted feature generation & selection for classification, Li et al. [5] provide an up-to-date review, including a breakdown of the most commonly used features and feature importance, as well as highlighting some of the current issues faced by the industry standard approaches using STA/LTA pickers which struggle at detecting events with low spatial and temporal separation. For recent approaches on deep learning-based classification approaches, including multi-class classification, please refer to [6].

The use of deep-learning to detect microseismic signals is becoming far more common and some of the main methods used are Convolutional Neural Networks (CNN)-based [6], [7], though other deep learning architectures have been used as well including a pseudo-Siamese implementation of EQ-Transformer performing template matching, which is then fed into P and S networks [8] and region-based CNNs (R-CNN) enabling capture of earthquake events in 1-D time series data across multiscale (time dilation) anchors [9]. In our previous work [10] we showed detection of earthquakes below mag-

[*]This work was partly supported by the EPSRC under grant agreement No EP/X01777X/1 and by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955422. We thank Dr. F. Provost for providing a catalogue of labelled events used in this paper.

nitude 1, down to 0.2 at distances of up to 50km from an uncatalogued site in Scotland with data gathered from a single 3 axis geophone located on glacial till and colluvium with depths up to 20m.

Self-Organizing Maps (SOM), are a type of network which learns via competitive learning. During competition with each other a wining neuron is decided. The winning neuron receives positive feedback when similar connections are made to close neurons and negative feedback when dissimilar neurons are nearby. Optimising this way results in a self organised group structure or map. [11]. They can be applied in a number of way within earth sciences. In [12], SOMs are used to track the changes in a mine. Initially STA/LTA is used to detect events and then from these events a number of features from both the time and frequency domain are extracted from seismic waveforms. Secondly they can be used in seismic facies classification [13] where changes in waveform between boundaries can be used to map a 2D slice and result in a clearly delineated map of the target area.

In this paper we show the effectiveness of our Siamese network at classification via comparison, and demonstrate the use of self-organising maps (SOMs) as a way of explaining events which may be mislabelled via the features generated by the Siamese networks final encoding layer. We propose using a convolution Siamese network to help to improve the cataloguing workflow. Siamese networks work using 2 (or more) branches, a single network is shown two inputs, an anchor and a test, which are labelled similar or not; the weights of network are then updated simultaneously from both branches. During testing a known signature is used as an anchor and a second test signature is then shown to the network. Both anchor and test pass through the same network (weights) and the differences between the encoded outputs from the network is calculated typically using euclidean distance. If the similarity score is within a threshold the two are assumed to be similar. We then demonstrate using SOM as a method to explore the reason behind a misclassifications and by choosing the nearest neighbourhood, suggest a new label.

2. METHODOLOGY

2.1. Dataset

2.1.1. Résif Dataset

For the validation of our model we use the Résif labelled dataset (available at [14]), which is taken from monitoring in a quarry in France and is accompanied by a catalogue compiled by Provost et al.[15]. The seismic records are acquired by two permanent arrays of the French Landslide Observatory OMIV (Observatoire Multi-disciplinaire des Instabilités de Versants) installed at the east and west sides of the Super-Sauze landslide (Southeast France) developed in weathered black marls [15]. Data is gathered by two sensor arrays (SZB & SZC), each with one 3D sensor and three 1D (Z-axis) sensors. The sampling rate is 250Hz and events are between 0.2 seconds to 105 seconds. The dataset spans three monitoring periods, 11 October to 19 November 2013, 10 November to 30 November 2014, and 9 June to 15 August 2015. For our experiments, we make use of the MT.SZC station as it has more complete data. Our model uses all three channels from the 3D sensor (sensor 0) from the SZC array.

Table 1. Résif catalogue events

Class	Total No.	Uncertain	Avg Dur.	Avg Max Freq.
Rockfall	402	0	16.16s	6.83Hz
Earthquake	382	0	13.06s	4.33Hz
Micro	234	17	2.25s	7.15Hz
Noise	340	0	10.37s	4.07Hz

As shown in Table 1, there are four types of labelled events: rockfall events showing distinctive impacts over several seconds, micro-earthquakes which are very short events less than 5 seconds, earthquake events which have triangular spectrogram components with reducing high frequency content ranging from 2-50 Hz, and finally, noise, natural or man-made, that can last tens of seconds and generally has higher frequency range between 0 and 100Hz and very distinct spectrogram. Note that the events in the Résif catalogue can be as low as 2Hz. See [15], [6], [16] for details about the four classes including waveform and STFT examples.

The second column in Table 1 shows the total number of events per class in the labelled Résif dataset. The labels were generated using an STA/LTA detector in the frequency domain and a supervised random forest algorithm and then expertly validated via visual inspection[15]. Even then, seismic data labelling can be challenging. Hence, in the catalogue there are a number of events which have been labelled as one class but contain a note suggesting there may be doubt, as shown in the third column labelled 'Uncertain'. This occurs most frequently within the micro-seismic class where 15 events are noted as being possible rockfalls and 2 possible earthquakes. There is a large imbalance of events in each monitoring period; for example, there are only 5 micro-quake events in 2015 and 339 rockfall events in 2014.

Within the earthquake class, there are also 9 teleseismic events between the 12^{th} and 25^{th} of October 2013, possibly aftershocks of the 2013 Bohl earthquake. The teleseismic events are longer duration than the other earthquake events (around 40 to 60 seconds), and hence are likely to lead to classification errors. We take all of the catalogue labels as ground truth during training and testing, and show examples of a label that differs from its class in Section 4.2, concluding that they are most likely mislabelled.

2.2. Model

The designed Siamese network architecture comprises of a two-branch feature extractor each processing one input,

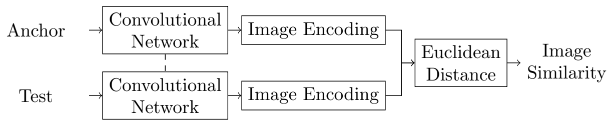


Fig. 1. Siamese network architecture

implementing feature learning transform τ via fully convolutional network. The comparison head implements a distance function. Among numerous distance functions used for Siamese networks, the Cosine Distance worked best.

$$\text{CosineDistance}(P, Q) = 1 - \frac{P \cdot Q}{\|P\|\|Q\|} \tag{1}$$

where P and Q are n-length vectors with entries p_i and q_i, respectively. The decision making head is a single dense neuron providing a soft-label in the range of 0 to 1, where two identical sequences would result in an output of 0 and highly dissimilar inputs would produce an output close to 1.

The convolutional layers and all but the last dense layer use 'relu' as their activation function; the final dense layer uses 'sigmoid'. The convolution layers are 2D due to the 3D based input (in our case 65×66 Short-time Fourier Transform (STFT) image output for each of the three channels). Dropout is used after the first two convolution layers and max pooling is used to reduce the size of the feature maps.

2.3. Clustering

Clustering is done using SOM, specifically from the miniSOM [17] python package. With the trained Siamese network the last Dense layer is set to the output, which allows the final encoding to be output rather than the result of the distance function. With this, SOM is used on the resulting encoding vectors to create a SOM map.

3. EXPERIMENTAL SETUP

The dataset is split into training & validation and testing. The training & validation events are sorted in chronological order and stratified. The test size was set to be the last 30% per class, and the test set was identical for all runs.

For the ablation study, multiple runs were performed for every configuration using sklearns' StratifiedKFold set to 5 folds. At the start of each fold the datasets (training, validation) were created using tensorflow.data.Dataset which then, were repeated, shuffled (with a set seed) and zipped. Initial training involved all of the training & validation data (70% of all data) being split into a 5-folds with 80% being used for training and 20% for validation.

Performance of the model was calculated using the test set, generating the distance matrix of the encoded test set and comparing against each other to obtain the performance across all 'anchors' of each class.

Training is completed using Python 3.10.10 using TensorFlow 2.10.1 on a system with a i7 10700K and RTX NVIDIA 2080 Ti. Throughput is around 46.5 STFT images per ms. With a dataset repetition of 10, a training epoch with validation takes around 15s.

4. RESULTS

4.1. Siamese Network Classification Performance

The classification results are shown in Table 2. The testing dataset consists of the last (chronologically) 30% of labels from each class. Table 2 shows how many events were correctly matched to the anchor for each class. For example, of the 117 rockfalls in the test set (as per Table 1), 107 rockfall events (True Positives) were correctly matched with similarity over the threshold and 116, 73 and 106, earthquake micro-quake and noise events, respectively, were correctly identified as dissimilar to the rockfall anchor (True Negatives). However, 10 labelled rockfall events were not matched with the anchor (False Negatives) and 1 micro-quake, 2 earthquake and 8 noise were incorrectly deemed similar to the rockfall anchor (False Positives for rockfall). Relative to class size micro-quake events perform the worst with 11 micro-to-micro comparisons not resulting in a high similarity score, and hence producing False Negatives. There are only a total of 55 misclassifications between one class and another, the majority involving noise. Besides noise, there are very few misclassifications compared to test size, only 11. Adequate anchor choice is therefore critical in being able to correctly detect all of events for each class.

4.2. Siamese Network Unsupervised Grouping

Manual picking and subsequent verification by experts is one of the slowest parts of labelling of seismic events, especially microseismic events with low signal-to-noise ratio. Using the encoded vectors from a trained network SOM can create an unsupervised map showing seismic events which are similar. It is possible to do the same with using on a Siamese network and comparing every event against each other but this may be unscaleable for extremely large datasets.

Table 2. Siamese network performance all test events confusion

	Test			
Anchor	rockfall	earthquake	micro-quake	noise
rockfall	107(10)	3	1	11
earthquake	1	114(5)	4	4
micro-quake	2	0	63(11)	12
noise	8	3	6	92(25)

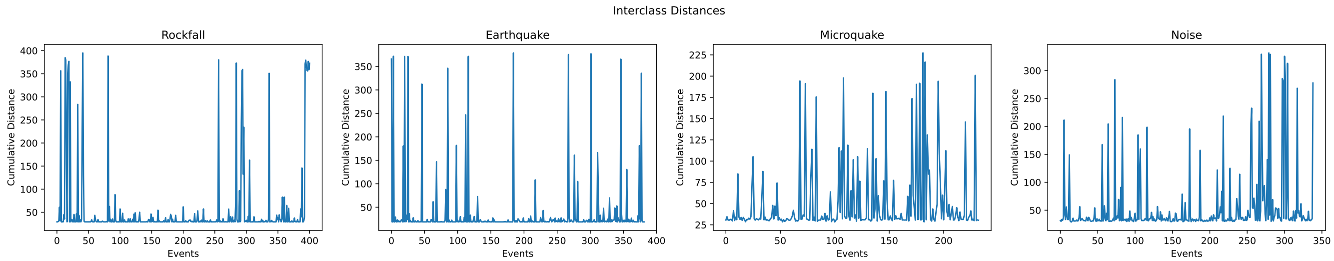
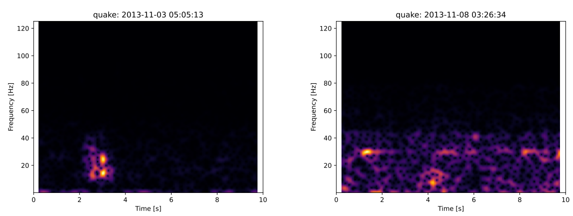


Fig. 2. noise interclass distance

Given an unseen and unlabelled dataset, by grouping similar events, the Siamese network enables clusters or “classes” to be formed with zero knowledge. That is, the network compares a randomly selected sample to all others and clusters all samples similar to the selected one together. Initially, this would be slow as large clusters are created, but quickly would reduce to more focused and smaller clusters.



Shown in Figure 2 is the inter-class distances between the different event types labelled in the Resif dataset. Low distance represents similar while high distance represents dissimilar events, it can be seen that from within the rockfall and earthquake class most events are of very low cumulative distance when compared against every other example in their respective classes, with a few exceptions which have distances close to the maximum (the maximum being the total number of events in the class, minus the event in question). For microquakes and noise there is more uncertainty in the events while many are still similar in some way to others within the class; in general the average distance is greater.

Self-organising maps offer another way of visually identifying outliers and helping to understand what class they might belong in as the distances are mapped into a 2D space. Figure 3 shows the distances between the different event types within the RESIF catalogue. The SOM map highlights helps to show the clusters of each class, as could be expected from the cumulative distances shown in Figure 2 the rockfall class has the least interference from other classes, but does have outliers that appear mainly in the noise and microquake section of the map. This makes sense as rockfall and microquake can appear highly similar if the SNR is very low or even if the initial rockfall impact is very large. Earthquakes have a few outliers which are mainly in the microquake class, again explainable by low magnitude earthquakes or distant where on the P-wave is visible above noise. As mentioned previously there is also the possibility that some microquakes were mislabelled and may in fact be rockfalls, the SOM would be able to help by highlighting those events which are closer in distance to the rockfall neighbourhood.

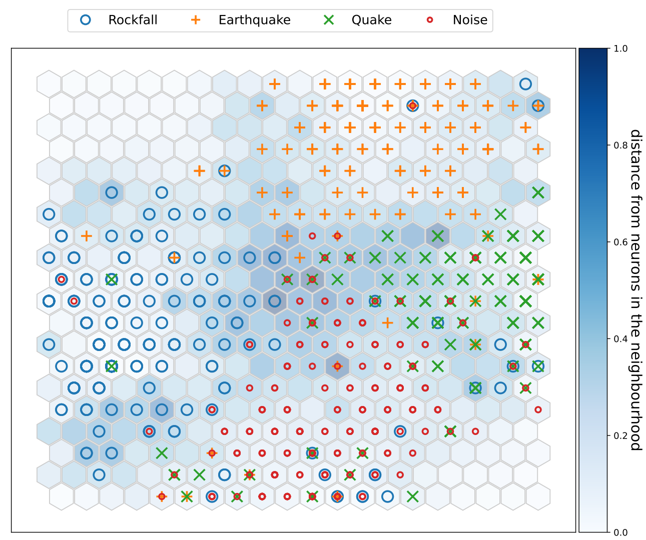


Fig. 3. Distance between events in the RESIF catalogue.

SOM is a helpful tool for experts to help visualise the output of the network and to understand where and why misclassifications have occurred.

5. CONCLUSION

We show the performance of our Siamese network in classification tasks, the ability to generate features and the usefulness of SOM to highlight more challenging labels which would require expert validation to improve labelled datasets generated by deep learning models.

6. REFERENCES

[1] Denis Anikiev, Claire Birnie, Umair bin Waheed, Tariq Alkhalifah, Chen Gu, Dirk J. Verschuur, and Leo Eisner, "Machine learning in microseismic monitoring," *Earth-Science Reviews*, vol. 239, pp. 104371, 2023.

[2] S. Grubas, S. Yaskevich, and A. Duchkov, "Localization of microseismic events using the physics-informed neural-network for traveltime computation," *82nd EAGE Annual Conference & Exhibition*, vol. 2021, no. 1, pp. 1–5, 2021.

[3] Marius Kriegerowski, Gesa M. Petersen, Hannes Vasyura-Bathke, and Matthias Ohrnberger, "A Deep Convolutional Neural Network for Localization of Clustered Earthquakes Based on Multistation Full Waveforms," *Seismological Research Letters*, vol. 90, no. 2A, pp. 510–516, 12 2018.

[4] Flora Giudicepietro, Antonietta M. Esposito, Laura Spina, Andrea Cannata, Daniele Morgavi, Lukas Layer, and Giovanni Macedonio, "Clustering of experimental seismo-acoustic events using self-organizing map (som)," *Frontiers in Earth Science*, vol. 8, 2021.

[5] Jiangfeng Li, Lina Stankovic, Stella Pytharouli, and Vladimir Stankovic, "Automated Platform for Microseismic Signal Analysis: Denoising, Detection, and Classification in Slope Stability Studies," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–11, 2020.

[6] Jiaxin Jiang, Vladimir Stankovic, Lina Stankovic, Emmanouil Parastatidis, and Stella Pytharouli, "Microseismic event classification with time-, frequency-, and wavelet-domain convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.

[7] Luca Trani, Giuliano Andrea Pagani, João Paulo Pereira Zanetti, Camille Chapeland, and Läslo Evers, "Deepquake — an application of cnn for seismo-acoustic event classification in the netherlands," *Computers & Geosciences*, vol. 159, pp. 104980, 2022.

[8] Zhuowei Xiao, Jian Wang, Chang Liu, Juan Li, Liang Zhao, and Zhenxing Yao, "Siamese earthquake transformer: A pair-input deep-learning model for earthquake detection and phase picking on a seismic array," *Journal of Geophysical Research: Solid Earth*, vol. 126, 2021.

[9] Yue Wu, Youzuo Lin, Zheng Zhou, David Chas Bolton, Ji Liu, and Paul Johnson, "DeepDetect: A Cascaded Region-Based Densely Connected Network for Seismic Event Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 62–75, 2019.

[10] David Murray, Lina Stankovic, Stella Pytharouli, and Vladimir Stankovic, "Semi-supervised seismic event detection using Siamese Networks," Tech. Rep. EGU23-14184, Copernicus Meetings, Feb. 2023, Conference Name: EGU23.

[11] Teuvo Kohonen, "Essentials of the self-organizing map," *Neural Networks*, vol. 37, pp. 52–65, Jan. 2013.

[12] S G Meyer, A M Reading, and A P Bassom, "The use of weighted self-organizing maps to interrogate large seismic data sets," *Geophysical Journal International*, vol. 231, no. 3, pp. 2156–2172, Sept. 2022.

[13] Shi'an Shen, Xinjian Wei, Xiaokai Wang, and Wenchao Chen, "Maximum correntropy self-organizing map for waveform classification," *GEOPHYSICS*, vol. 88, no. 3, pp. IM51–IM60, May 2023.

[14] French Landslide Observatory Seismological Datacenter RESIF, "Observatoire multi-disciplinaire des instabilites de versants (omiv)," 2023.

[15] F. Provost, C. Hibert, and J.-P. Malet, "Automatic classification of endogenous landslide seismicity using the random forest supervised classifier," *Geophysical Research Letters*, vol. 44, no. 1, pp. 113–120, 2017.

[16] Jiangfeng Li, Lina Stankovic, Vladimir Stankovic, Stella Pytharouli, Cheng Yang, and Qingjiang Shi, "Graph-based feature weight optimisation and classification of continuous seismic sensor array recordings," *Sensors*, vol. 23, no. 1, 2023.

[17] Giuseppe Vettigli, "Minisom: minimalistic and numpy-based implementation of the self organizing map," 2018.