GRAPH-BASED MICRO-SEISMIC SIGNAL CLASSIFICATION WITH AN OPTIMISED FEATURE SPACE

J. Li[*], *C. Yang*[†], *V. Stankovic*[*], *L. Stankovic*[*], *S. Pytharouli*[*]

[*] University of Strathclyde, Glasgow, United Kingdom
[†] York University, Toronto, Canada

ABSTRACT

Classification of seismic events detected from seismic recordings has been gaining popularity in interpretation of subsurface processes, e.g., volcanic systems, earthquake activity, induced seismicity and slope stability, in particular landslides.   However, due to the variability of signal representation for different classes in the temporal and spectral space, a large feature space to characterise the uniqueness of a particular type of event is used for classifying seismic signals. The consequence is additional complexity on the classifier and overfitting. So far, there has been little attempt to address dimensionality reduction via feature selection. In this paper, we propose an iterative, alternating graph feature and classifier learning method for micro-seismic signals via graph Laplacian regularization and normalized graph Laplacian regularization.   Using recorded micro-seismic events from an active landslide, we demonstrate improved classification accuracy with a relatively small feature space compared to state of the art.

Index terms— Micro-seismic, graph feature learning, GLR, norm-GLR

1. INTRODUCTION

Microseismic monitoring has been increasingly used to illuminate subsurface processes, such as landslides, due to its ability to detect small seismic waves generated by soil movement and/or brittle behaviour of rock [1].   The advantage of microseismic against traditional geodetic and geotechnical monitoring for slope stability is that seismometers can be deployed outside the unstable soil mass, offering long-term continuous monitoring over a large area [2]. However, recordings are often contaminated by signals generated by other sources (wind, rainfall, human activities, animals, etc.), requiring robust and sophisticated methods to characterise the signals that represent slope instability [1].

To extract useful information from micro-seismic recordings, it is necessary to denoise measurements, detect potential micro-seismic events, construct and select most discriminative features, classify the events into classes based on the type of the micro-seismic activity, and finally localize the events.

In this paper, we leverage on developed signal pre-processing and detection methods [2], [3], [4] and focus on signal feature selection and classification. Commonly, feature selection is performed manually, based on expert knowledge to select the most discriminative feature vector [5] or estimating the variable importance [6]. Following recent advances in graph signal processing (GSP) [7], and more specifically graph regularization-based classification [8] and graph learning [9], we develop and evaluate suitability of GSP-based methods for automatic feature selection and classification. We show that the integrated iterative feature selection, extraction and classification method outperforms state-of-the-art feature selection and classification approaches, reducing the overall complexity.

2. BACKGROUND

The four classes of micro-seismic signals of interest to this study are [2]: (i) Earthquakes: these are potential landslide triggers, usually with medium to long duration, their frequency and waveforms are affected by the source distance, which makes classification challenging; (ii) Slidequakes: can be seen as earthquakes with a short duration; however, there is no clear dominant frequency; (iii) Various tremors: observed as episodic tremor and slip (ETS) signals, harmonic, rockfall and multi-events signals; (iv) Calibration shots: artificial signals generated to calibrate the seismic monitoring system, derive velocity models and test localization algorithms.

Micro-seismic signal detection and classification usually starts by detecting candidate events of micro-seismic activity, which is traditionally performed using statistical approaches (e.g., STA/LTA or Akaike Information Criterion, frequency analysis [3]), or combing the two [2]. After potential events are identified, feature construction and selection take place. The current trend is focused on mostly feature construction, resulting in a very large set of features (e.g., 102 in [10]) including signals of ascending and descending duration, signal duration, signal

envelop skewness and kurtosis, and energy of different frequency bands [5], [6]. However, high number of features could overfit classification models and result in non-optimal performance. This motivates feature selection to select the most discriminative features from a subset of the previously constructed features. While seismic signal feature selection research is an emerging field, initial steps include using expert knowledge to reduce the feature vector from 63 to 14 in [5], and using Variable Importance estimation to reduce the feature set to 58 from 71 in [6].

2.1. Graph signal processing (GSP)

Graphs are commonly used data representation forms for irregular data domains such as social networks, biological networks, point cloud data, sensor array displacement, etc., by capturing inter-dependency between collected data via weighted vertices [7].

Let $\mathcal{G} = (\mathbf{V}, \mathbf{A})$ be an undirected graph, where $\mathbf{V}$ is the set of vertices and $\mathbf{A}$ is the weighted adjacency matrix. By assigning each data sample $x_i \in R^n$ to one graph vertex, $v_i \in \mathbf{V}$, and seeing classification labels as a piece-wise signal on graphs, one can restore the missing labels and perform classification by smoothing the graph signal using graph Laplacian regularization [7].

Indeed, let $f_k(i)$ be the k-th feature of data sample x_i associated to vertex v_i, and let $F_k = (f_k(i) - f_k(j))^2$. Then we define the adjacency matrix $\mathbf{A}$ with its (i,j)-th entry:

$$a_{i,j} = exp\left\{-\sum_{k=1}^{K} c_k \frac{F_k(i,j)}{\sigma_k^2}\right\}, \qquad (1)$$

where σ_k is the scaling factor, c_k weights the importance of the k-feature, and K is the total number of features.

The graph Laplacian matrix is then defined as $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is the diagonal matrix, given by $d_{i,i} = \Sigma_j a_{i,j}$, and its normalized form is $\mathbf{L}_{norm} = \mathbf{D}^{(-1/2)}\mathbf{L}\mathbf{D}^{(-1/2)}$.

For a binary (semi-)supervised classification problem, where all N data samples x_i need to be classified into one of two pre-defined classes, based on the training set of features $f_k, \ldots, f_n$, $n < N$, a graph signal representing classification labels is defined as:

$$s_i = \begin{cases} +1, & \text{if } x_i \text{ belongs to the Class and } i \leq n \\ -1, & \text{if } x_i \text{ does not belong to the Class and } i \leq n \\ 0, & \text{for } n < i \leq N. \end{cases}$$
$$(2)$$

Then, one can perform effective classification, competitive to advanced classification approaches when the training set is unreliable or very small [7], by minimizing the (normalized) graph Laplacian regularisation (GLR) term given by:

$$s^* = \operatorname*{arg\,min}_{s \in R}(s^T \mathcal{L} s) \qquad (3)$$

where $\mathcal{L}$ is either graph Laplacian or normalised graph Laplacian.

Graph-based feature weight learning so far concentrates on diagonal element optimization [9], [11], which assumes features are independent while full matrix optimization [12], [13] takes into account cross-correlation among features. However, the search space of the optimal solution in [13] is not large enough and [12] uses a closed-form solution without optimization. In this paper, we focus on diagonal element optimization only. As a benchmark, we implement the seminal work on graph feature learning [11], that adopts gradient descent to determine the diagonal entries while minimizing the average entropy of the unknown labels.

3. METHODOLOGY

Based on prior micro-seismic event characterisation [2] as well as time-series acoustic signal classification [10], we narrowed down and constructed a set of 48 features, shown in Table 1, to characterise micro-seismic events arising from landslides. The 48 candidate features include spectrum features (feature 1-15), temporal features (feature 16-42), and cepstrum features (feature 43-48). The task is to select an optimum subset of the 48 candidate features automatically and classify the corresponding events into one of four classes (see Section 2).

Table 1: Candidate Features: $y(t)$ is the seismic signal in time domain, $p(t)$ is the power signal, $e(t)$ is the signal envelope, $q(v)$ is the frequency domain signal and $c(v)$ is the cepstrum.

1.Mean of $q(v)$	17.Root mean square $y(t)$	33.Rate of attack $y(t)$
2.Max of $q(v)$	18.Standard deviation $y(t)$	34.Rate of attack $e(t)$
3.Median of $q(v)$	19.Max value of $y(t)$	35.Rate of decay $y(t)$
4.Variance of $q(v)$	20.Average power of $y(t)$	36.Rate of decay $e(t)$
5.Max envelop PSD of $q(v)$	21.Skewness of $y(t)$	37.Ratio1=30/31
6.Frequency 95%bandwidth	22.Skewness of $p(t)$	38.Ratio2=30/32
7.Frequency 50%bandwidth	23.Median value of $y(t)$	39.Ratio3=30/std $e(t)$
8.Dominate frequency	24.Kurtosis of $y(t)$	40.Ratio4=$(t_{max}-1)/(N-t_{max})$
9.Median normal $q(v)$	25.Kurtosis of $p(t)$	41.Onset of $y(t)$
10.Variance normal $q(v)$	26.Entropy of $p(t)$	42.Offset of $y(t)$
11.Int-ratio [14]	27.Entropy of $y(t)$	43.Std of $c(v)$
12.Number of peaks of $q(v)$	28.Kurtosis of $e(t)$	44.Mean of $c(v)$
13.Gamma 1 [14]	29.Skewness of $e(t)$	45.Skewness of $c(v)$
14.Gamma 2 [14]	30.Max of $e(t)$	46.Kurtosis of $c(v)$
15.Gamma 3 [14]	31.Meam of $e(t)$	47.Dominate Frequency of $c(v)$
16.Duration $y(t)$	32.Median of $e(t)$	48.Max of $c(v)$

Alg. 1 represents the proposed iterative alternating binary graph classification and feature learning for micro-seismic signal classification. Feature selection is performed iteratively between classification and feature weight update during graph learning. The algorithm takes as input the entire set of candidate features, $f_k(i)$, and the set of micro-seismic events to be classified x_i, where the first n events are labelled and used for training. The algorithm processes one class at the time based on one-against-all strategy converting this multi-class classification problem into multiple binary classification tasks. Eq. (2) is used to form the initial graph signal, where

Algorithm 1 Alternating Classifier and Graph Learning

Input:

    Candidate Feature matrix (Table 1)

    Initial graph signal s_i (2)

    Graph weight scaling factor σ_k, see (1)

    $tol\_set = tol\_pg\_set = \mu$

    Constraint C, such that $\sum_k c_k \in \left[0, C\right]$

Output: The predicted labels $s^{predicted}$

 1: **Initialization**: c_k identity vector, $tol = 1$; $i = 1$

 2: **while** $tol > tol\_set$ **do**

 3:    $\mathbf{A} \leftarrow$ (1) and find $\mathcal{L}$

 4:    $s^* \leftarrow$ (3); $tol\_pg = 1$, $j = 1$

 5:    **while** $tol\_pg > tol\_pg\_set$ **do**

 6:        $c_k^* \leftarrow proximal\ gradient\ descent$ [9]

 7:        update $\mathbf{A} \leftarrow$ (1) and find $\mathcal{L}$ using c_k^*

 8:        $obj\_pgvalue(j) \leftarrow$ using (4)

 9:        $s^* \leftarrow$ (3); $j = j + 1$

10:        $tol\_pg = obj\_pgvalue(j) - obj\_pgvalue(j-1)$

11:    $obj_{value}(i) = \min\limits_{s \in R}(s^T \mathcal{L} s)$

12:    $tol = obj_{value}(i) - obj_{value}(i-1)$; $i = i + 1$

13: **return** $s^{predicted}$

Class denotes the micro-seismic class to be considered for each binary classification.

The hyper-parameters, σ_k, C, and the stopping criteria μ are heuristically set. Note that $c_k \geq 0$ represents the weight given to feature f_k and is set to zero if f_k is not used at all.

Based on [9], Alg. 1 alternates between learning the features, c_k^*, in Line 6 for fixed classifier (s^*) and learning the classifier s^* for a fixed set of features c_k^*. The objective function for graph learning in Line 8 is given by [9]:

$$min \sum_{i,j} exp \left\{ -\sum_{k=1}^{K} c_k F_k(i,j) \right\} d_{i,j} + I(\mathbf{c}), \quad (4)$$

where $I(\mathbf{c})$ is an indicator function, that returns 0 if all elements of $\mathbf{c} = [c_1 \ldots c_K]$ are in the range $[0, C]$, or ∞, otherwise. The algorithm iterates until the improvement is above the threshold μ. Finally, as the output s^* is a real number, we take the sign operation as the final label: $\widetilde{s}^* = sign(s^*)$. See [9] for the complexity and convergence analysis.

4. EXPERIMENTAL SETUP AND RESULTS

We evaluate the proposed classifier and graph learning on a landslide site in Super-Sauze with continuously recorded micro-seismic data for 58 days in 2010, containing four event types or classes: earthquake, slidequake, tremors, and calibration shot [2].

Firstly, similarly to [4], we perform bandpass filtering and Neyman-Pearson detection resulting in 223 detected events (56 earthquakes, 58 slidequakes, 37 tremors, 11 calibration shots, and 12 undefined events) plus 49 false alarms. Note that [4] showed that Neyman-Pearson outperforms STA/LTA, usually used for event detection. After feature construction, we implement the SMOTE algorithm to generate synthetic feature samples based on all 48 features to balance the training set [15]. We split the entire set into 30% used for testing and 70% used for training, where the testing set does not contain any synthetic features. In Alg. 1 we set $\mu = 0.01$ and $C = 48$ (the total number of features).

As benchmark we use the feature selection method of [16] (Inf-FS) (as used in [4]) and classifiers - Support Vector Machine (SVM) and Random Forest (RF), as used in [10] and [6], respectively. Inf-FS is a filter-based unsupervised feature ranking method, that outputs features sorted by their importance. Then, cross-validation is utilized to pick the optimized subset of features with best classification performance. Finally, SVM or RF classifiers are used for classification based on the selected features. In addition, we also implemented the harmonic function-based graph feature learning method of [11] for benchmarking.
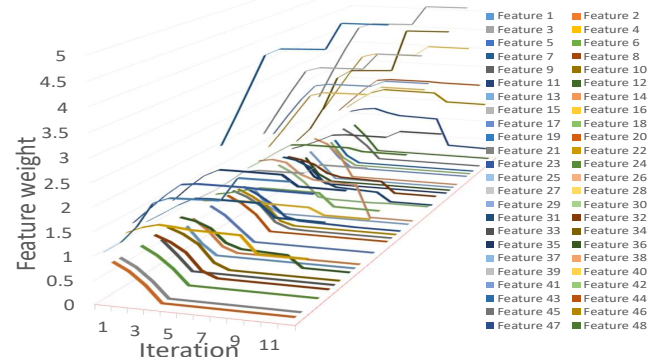
Fig. 1. Variation of feature weights, for all 48 features, with iteration number.

Figure 1 shows how the feature weights for all 48 starting features vary across iterations. We can observe the variations in feature weights between iterations until convergence is reached (the curves become flat) around the tenth iteration, even earlier for most features. This is therefore the number of iterations we use in our experiments, whose results are shown in Table 2. Furthermore, we also observe only 22 features with non-zero weight once convergence is reached, showing dimensionality reduction of over 50%.

Table 2 shows the classification performance expressed as F-score, where 1 is the best performance and 0 the worst. We compare the non-iterative Inf-FS feature selection followed by the ML classifier (SVM or RF), as well as the method of [11], with our proposed iterative alternating graph learning with two graph-based classifiers. We can observe that the proposed graph-based alternating classifiers and feature learning, for all 4 micro-seismic

event classes, provides the most accurate classification result. The improvement is especially significant for the Slidequake class. The reason that benchmark [11] performs slightly worse than our method may potentially be that it does not promote smoothness within the label signal.

Table 2: Classification Result (F1-score)

Feature Selection	Classifier	Earthquake	Slidequake	Tremors	Calibration Shot
Inf-FS	SVM	0.9	0.84	0.92	0.8
	RF	0.9	0.88	0.91	0.9
[11]	[11]	0.91	0.91	0.88	0.8
Graph Learning	GLR	0.94	**0.95**	0.93	0.85
	norm GLR	**0.95**	0.93	**0.95**	**0.92**

5. CONCLUSION

We propose an iterative alternating graph feature and classifier learning approach to tackle the emerging problem of micro-seismic event classification. We show that by tackling feature selection and classification jointly as in the proposed approach, we can consistently improve classification performance, with a smaller set of features, of 4 types of micro-seismic events compared to state-of-the-art seismic event classification methods. Future work will focus on optimizing the underlying graph, i.e., adjacency matrix and considering the case of labelling errors in the training set.

6. ACKNOWLEDGEMENT

7. REFERENCES

[1] J. Whiteley, J. Chambers, S. Uhlemann, P. Wilkinson, and J. Kendall, "Geophysical monitoring of moisture-induced landslides: a review," *Rev. Geophys*, vol. 57, no. 1, pp. 106–145, 2019.

[2] N. Vouillamoz, S. Rothmund, and M. Joswig, "Characterizing the complexity of microseismic signals at slow-moving clay-rich debris slides: the super-sauze (southeastern france) and pechgrabe (upper austria) case studies.," *Earth Surface Dynamics*, vol. 6, no. 2, 2018.

[3] A. C. To, J. R. Moore, and S. D. Glaser, "Wavelet denoising techniques with applications to experimental geophysical data," *Signal Processing*, vol. 89, no. 2, pp. 144–160, 2009.

[4] J. Li, L. Stankovic, S. Pytharouli, and V. Stankovic, "Automated platform for microseismic signal analysis: Denoising, detection and classification in slope stability studies," *unpublished*, submitted Dec2019.

[5] M. Curilem, A. Cuevas, R. Soto, F. Huenupan, C. Martin, M. S. Khan, F. Gil, C. Cardona, and L. Franco, "Classification of volcanic seismic events: An expert knowledge analysis for feature selection," *ICPRS*, 2017.

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

[7] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[8] M. Belkin and P. Niyogi, "Using manifold stucture for partially labeled classification," in *NeurIPS*, 2003.

[9] C. Yang, G. Cheung, and V. Stankovic, "Alternating binary classifier and graph learning from partial labels," in *(APSIPA ASC)*, IEEE, 2018.

[10] M. Malfante, M. Dalla Mura, J.-P. Métaxian, J. I. Mars, O. Macedo, and A. Inza, "Machine learning for volcano-seismic signals: Challenges and perspectives," *IEEE Signal Processing Mag*, vol. 35, no. 2, pp. 20–30, 2018.

[11] X. Zhu, Z. Ghahramani, and J. D. Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," in *ICML*, 2003.

[12] P. Zadeh, R. Hosseini, and S. Sra, "Geometric mean metric learning," in *J. Mach. Learn. Res*, 2016.

[13] W. Hu, X. Gao, G. Cheung, and Z. Guo, "Feature graph learning for 3d point cloud denoising," *arXiv preprint arXiv:1907.09138*, 2019.

[14] A. Maggi and et al, "Implementation of a multistation approach for automated event classification at piton de la fournaise volcano," *Seismol Res Lett*, vol. 88, no. 3, pp. 878–891, 2017.

[15] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *J. Artif. Intell. Res*, vol. 16, pp. 321–357, 2002.

[16] G. Roffo, "Feature selection library (matlab toolbox)," *arXiv preprint arXiv:1607.01327*, 2016.