Active Learning Applications to Acoustic Emission Data

L. Bull[1a], K. Worden[2], N. Dervilis[3]

[123] Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield,

Mappin Street, Sheffield S1 3JD, UK

[a] lbull1@sheffield.ac.uk

Abstract. The intention of this work is to demonstrate a strategy for using Active Learning algorithms to investigate Acoustic Emission (AE) data. An informative chain of artificial intelligence tools will allow a semi-supervised interaction between meaningful clusters of data, to help select the most informative samples for efficient training.

Introduction

Active-learning, or query-learning, is a semi-supervised machine learning approach. Its main premise is that a learner can perform with increased efficiency, using less training data, if it can select the data from which it learns [1]. This method becomes particularly relevant when data may be abundant but labels are difficult to obtain. Consequently, this learning philosophy is especially appropriate to the remit of AE monitoring.

Problem Statement. Whilst a large supply of AE data may be available, the exact information describing the emissions is often in short supply; whether it's simply the source location, or the mechanism that caused it – crack initiation, crack-growth or other frictional effects. The purpose of this work is to demonstrate the potential advantages of applying active-learning algorithms to classify AE data, along with any other necessary machine learning and signal processing techniques.

Active learning. A wide range of active learning algorithms exist within the remit of machine learning. Some examples include the 'Active Support Vector Machine' (SVM_{active}) [2] and 'Manifold Adaptive Experimental Design' (MAED) [3]. Each employ various 'querying' methods to determine which unlabelled data carry the most information; some measures include the distance from the decision boundary (SVM_{active}) or the average squared predicted error (MAED). The fundamental steps behind active learning tools are summarised below:

1. Provided unlabelled input data.
2. By some querying component, determine which of the unlabelled instances to query.
3. Provide labels for these data.
4. The learner is then trained these informed data.

A visual representation of the active learning steps is shown in Fig.1; this represents a three class problem, using toy data generated in a two-dimensional feature space. Fig.1 illustrates the smaller, more informed data sets that active learning techniques can provide.

Engineering application

For context, an example of real acoustic emissions data [4] is provided in Fig.2; four traditional AE features – rise time, peak amplitude, duration and ringdown count – have been projected on to two dimensions through principal component analysis. It can be observed from these real data that the application of active learning tools would be advantageous:

- A random sample of data for training might result in a poor representation of all clusters.
- Active learning tools will allow a querying regime to establish a small but informative training set to be labelled.

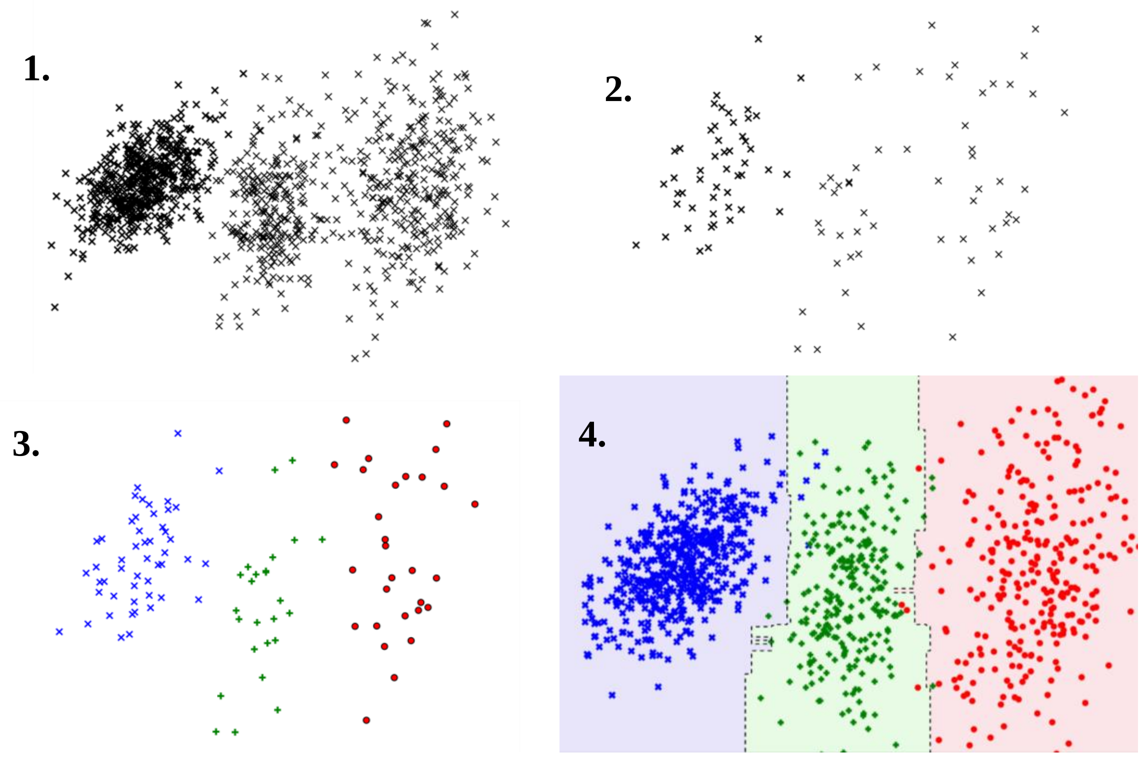

Figure.1 Visual representation of active learning steps

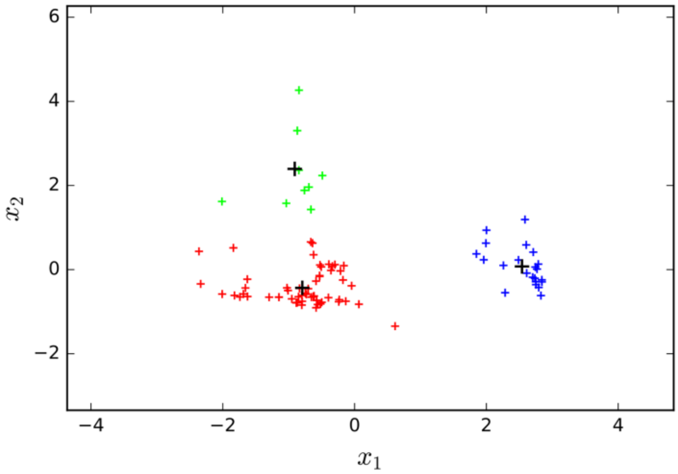

Figure.2 Acoustic Emission data in two-dimensional feature space

Acknowledgments

Thank you to Karen Holford and Rhys Pullin at Cardiff University for providing the AE data.

References

[1] B. Settles: *Active learning literature survey*. University of Wisconsin, Madison, vol. 52 (2010), p. 11.

[2] S. Tong, and K. Daphne: *Support vector machine active learning with applications to text classification.* J. of machine learning research vol.2 (2001) p. 45-66.

[3] Cai, Deng, and H Xiaofei: *Manifold adaptive experimental design for text categorization* IEEE Transactions on Knowledge and Data Eng vol.24 (2012) p. 707-719.

[4] G. Manson, K. Worden, et al: *Visualisation and dimension reduction of acoustic emission data for damage detection* J. of Intelligent Material Systems and Structures vol. 12 (2001) pp. 529-36.