A variational Bayesian clustering approach to acoustic emission interpretation including soft labels

Martin Mbarga Nkogo[1], Emmanuel Ramasso[1][0000−0002−1744−7453], Patrice Le Moal[1][0000−0001−8431−6210], and Gilles Bourbon[1][0000−0003−1580−9876]

Univ. Bourgogne Franche-Comté, FEMTO-ST Institute,
UFC/CNRS/ENSMM/UTBM, Department of Applied Mechanics, F-25000 Besançon

Abstract. We investigate Gaussian Mixture Models (GMM) with uncertain parameters to evaluate whether this model can help in interpreting acoustic emission data used in non-destructive testing. This model, called VBGMM (variational Bayesian GMM) allows the end-user to automatically determine the number of clusters which makes it relevant for this type of application where clusters are related to damages. In this work, we modify the training procedure to include prior knowledge about clusters. Experiments are made on a recently published benchmark, ORION-AE, that aims at estimating the tightening levels in a bolted structure under vibrations. Preliminary results of the VBGMM with soft priors (VBGMM-SOFT) show good improvement over the standard VBGMM.

Keywords: Clustering · Soft Labels · Acoustic Emission · Approximate inference · Structural Health Monitoring · Loosening of bolted joints · ORION-AE benchmark.

1 Introduction

The idea of using soft labels in clustering introduced in [21, 20, 4] for the Gaussian Mixture Model (GMM) and more general models [7] was motivated by the possibility to introduce the available knowledge on the components of a mixture model used to generate each observation. By using belief functions, the end-user can encode imprecision and uncertainty on the labels used in training and inference. Several studies demonstrated that the use of soft labels (using belief functions and probability theories) in various clustering and classification methods improves not only the global performance [6], but also the interpretation of clusters by providing insights about the decision frontiers [12], and the robustness against mislabelling [4, 15].

The present study aims at investigating how the Variational Bayesian GMM [2] (VBGMM) behaves when soft labels are introduced. This work is motivated by an application related to Structural Health Monitoring (SHM) based on the Acoustic Emission (AE) non-destructive technique.

The AE technique relies on permanently attached piezoelectric sensors glued on the surface of a material. Under loading, damage occurring within the material releases energy, and a part of it takes the form a high frequency (possibly between 20 kHz and 2 MHz) transient elastic wave propagating on the surface, converted into a voltage signal by the sensors. The technique is widely used in industry to detect anomalies, such as in civil infrastructures and aeronautics [19, 10, 9]. The advantage of this technique is the high sensivity of the available sensors which allows to get a lot of details about the damages. With a sampling frequency generally around 5 MHz on multiple sensors, the AE technique provides from thousands to millions of signals during mechanical tests.

There is no physics-based model able to interpret all the collected transient signals due to the several difficulties, in particular the unknowns about the influence of damages on the content of transient elastic waves. Therefore, the main methodologies to interpret AE data are mainly based on clustering, where the clusters are analyzed *a posteriori* to assess their relevance for a given application. The most widely used algorithms are the K-means [3], the fuzzy C-means (FCM) [11], the Gustafson-Kessel (GK) algorithm [13] and Gaussian Mixture Models (GMM) [18].

In previous studies, authors made use of clustering validity indices to estimate the number of clusters [22, 23], which is of paramount interest in Material Science and for AE users because it indicates the number of AE sources which are related to damages. One of the advantage of the VBGMM is its ability to automatically estimate the number of clusters and can thus be of interest for interpreting AE data. By allowing the introduction of soft labels in this model, we expect to improve the results even with small amount of prior.

Section 2 presents how to introduce soft priors in a VBGMM. Our first results are illustrated in Section 3 on a benchmark recently proposed for AE data clustering and classification.

2 Use of soft labels in a variational Bayesian GMM

2.1 Directed acyclic graph

A Bayesian Gaussian Mixture Model (GMM) is represented by the directed graph in Figure 1 where $\mathbf{y}_n$ is the value taken by an observed variable $\mathbf{Y}$, made of D-dimensional features in $\Re^D$ and $n = 1 \dots N$ the number of feature vectors, $\mathbf{x}_n$ is the value taken by a latent variable $\mathbf{X}$. Like in a standard GMM, $\boldsymbol{\pi}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ are the mixing proportions, means and precision (inverse of covariance).

In a Bayesian GMM, the three last parameters are uncertain and are considered as random variables to which the end-user assigns a prior distribution: A Dirichlet prior over $\boldsymbol{\pi}$, and an independent Gaussian-Wishart prior [2, Chap. 2] on $(\boldsymbol{\mu} \mid \boldsymbol{\Lambda})$ (mean and precision) of each Gaussian component [2, Chap. 10]. These particular priors are said conjugate because the posterior distributions have the same functional form as the priors through Bayes rule. The learning process consists in estimating the distribution over the uncertain parameters using a Bayesian Expectation-Maximization algorithm [1].

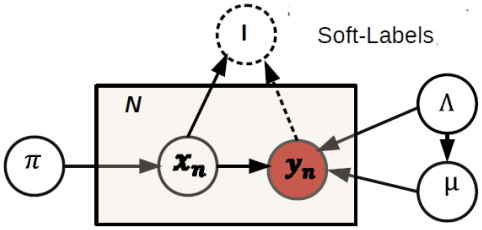


Fig. 1: Directed graph of a VBGMM-SOFT with prior on the latent variables $\mathbf{x}$.

In this graphical model, the prior (expert judgments) is represented by variable $\mathbf{I}$. In this "data-driven Bayesian network with expert judgments" [5], the values I_n taken by this variable are dependent on the values of $\mathbf{x}_n$ and may be dependent on $\mathbf{y}_n$, for example when the end-user tunes I_n according to the observed values and what he expects on $\mathbf{x}_n$.

2.2 Learning problem under *pl*

Solving the learning problem relies on a process detailed in [1] and [2, Chap. 10] and consists in maximizing the lower bound of the log-marginal probability of the data $p(\mathbf{X})$ subject to a factorization constraint. Indeed, a solution to this learning problem, called Variational Bayesian Expectation-Maximization, assumes a factorization between, on one side, a first factor that is the distribution over the parameters ($\boldsymbol{\pi}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$) and, on the other side, a second factor which is the distribution over the latent variables $\mathbf{x}_n, n = 1 \dots N$. Then, for one of those factors, we compute the *expectation of the logarithm of the joint distribution over all hidden and visible variables and then take the expectation with respect to the other factor*. The process is general can be applied to any mixture model.

The expectations are not trivial, the reader can refer to [2, Chap. 10] for more details. We here remind some of the results helpful to understand how to introduce the prior on $\mathbf{x}_n$. The other equations remain the same.

The first step is to compute the joint distribution on all variables in this model. Following Figure 1, it is given by:

$$p(\mathbf{X}, \mathbf{Y}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \mathbf{I}) = p(\mathbf{X} \mid \boldsymbol{\pi}) p(\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) p(\boldsymbol{\pi}) p(\boldsymbol{\mu} \mid \boldsymbol{\Lambda}) p(\boldsymbol{\Lambda}) p(\mathbf{l}|\mathbf{X}) \tag{1}$$

where

$$p(l_n|x_{nk} = 1) = pl_{nk} \tag{2}$$

represents the prior on the latent variable for the n-th feature vector and the k-th component in the mixture. This prior is represented by a plausibility contour function for each feature vector, generated from a belief mass over the set of K

components. The way of introducing the prior using an auxiliary variable (here $\mathbf{I}$) was proposed in [8].

Using the cognitive independence assumption [7], we can write

$$p(\mathbf{l}|\mathbf{X}) = \prod_{n=1}^{N} \prod_{k=1}^{K} [pl_{nk}]^{x_{nk}} \tag{3}$$

Introducing the prior on the latent variables modifies the expression of the expectation of x_{nk}, and does not change the Maximization step. Only the expectation on the latent variable (the values taken by x_{nk}) are modified by the plausibilities as follows:

$$\mathbb{E}[x_{nk}] \propto pl_{nk}\tilde{\pi}_k|\tilde{\boldsymbol{\Lambda}}_k|^{1/2} \exp\left[-\frac{D}{2\beta_k} - \frac{\nu_k}{2}\left(\mathbf{y}_n - \mathbf{m}_k\right)^{\mathrm{T}} \mathbf{W}_k \left(\mathbf{y}_n - \mathbf{m}_k\right)\right] \tag{4}$$

with

$$\log \tilde{\pi}_k = \psi\left(\alpha_k\right) - \psi\left(\sum_{k=1}^{K} \alpha_k\right) \tag{5a}$$

$$\log|\tilde{\boldsymbol{\Lambda}}_k| = \sum_{i=1}^{D} \psi\left(\frac{\nu_k + 1 - i}{2}\right) + D\log 2 + \log|\mathbf{W}_k| \tag{5b}$$

and $\psi(a) = \dfrac{d}{da}\Gamma(a)$ is the digamma function, $(\alpha_k, \beta_k, \nu_k\mathbf{W}_k, \mathbf{m}_k)$ are the parameters of the Dirichlet and Gaussian-Wishart distributions [2, Chap. 2,10]. The interesting point with this result is that Eq. 6 boils down to the expression found for non Bayesian GMM in [4] when uncertainty on parameters tends to 0.

2.3 Algorithm and Automatic Relevance Determination

The algorithm starts with random initial values and updates the parameters iteratively until the maximum number of iterations is reached (2000) or the evolution of the likelihood becomes less than 10^{-8}. The general algorithm is provided in Alg. 1.

One of the interests of this algorithm lies in the way some of the components vanish during learning. And this is possible without numerical instabilities as observed in standard GMM when K, for example, is too large. This phenomenon qualified as "Automatic Relevance Determination" allows the end-user to actually set the number of clusters to a large value, and, after convergence, some of the clusters can be removed due to the fact that several parameters tend to their prior. In particular, the *expected values of the mixing coefficients* $\mathbb{E}[\boldsymbol{\pi}]$ *in the posterior distribution* tend to $\alpha_0/(K\alpha_0 + N)$ for K clusters and N data points. In this expression, α_0 (the same for all components, set to 1 in experiments) is the parameter of the Dirichlet distribution over the mixing coefficients $\boldsymbol{\pi}$, which is the effective prior number of observations associated with each component of the mixture. Therefore, a cluster made of $\mathbb{E}[\pi_k]$ elements can be removed.

Algorithm 1 General algorithm of VBGMM-SOFT.

1: Generate plausibilities with ρ provided by end-user
2: **while** max_iterations not reached and evolution of likelihood above threshold **do**
3: E-Step: Compute $\log q^\star(\mathbf{X}) = \mathbb{E}_{\boldsymbol{\pi},\boldsymbol{\mu},\boldsymbol{\Lambda}}[\log p(\mathbf{X}, \mathbf{Y}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \mathbf{I})] + c$ (c is a constant)
4: For this step, use the same equations as in [2] except for $\mathbb{E}[x_{nk}]$ where Equation 5 should be used instead.
5: M-Step: Compute $\log q^\star(\boldsymbol{\pi})$, $\log q^\star(\boldsymbol{\mu})$ and $\log q^\star(\boldsymbol{\Lambda})$ by taking the expectation of Eq. 1 on, respectively, $\boldsymbol{\pi}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$, with respect to $\mathbf{X}$ (using the results of the E-Step).
6: Compute the likelihood (expression given in [2, p. 481] and using Equation 5).
7: **end while**

3 First results and first conclusion

This section presents preliminary results on the capacity of the VBGMM with soft labels to provide relevant clusters. For that we used a benchmark, called ORION-AE, obtained from a real system, and described in [17].

3.1 Data set description

The system is composed of a two metallic plates jointed by three bolts and was designed to reproduce the loosening phenomenon observed on structures made of assemblies, in particular when submitted to vibrations. One of the bolts was untightened manually to simulate the loosening. The lower plate was submitted to a 120 Hz harmonic force by means of a shaker. Seven levels of tightening were considered, and for each level, an acoustic emission sensor recorded the transients liberated during the test. Each tightening level was maintained during 10 seconds. The test was repeated 5 times, leading to 5 datatsets, each made of 7 seven classes with 70 seconds of data for different sensors (in this paper only the second sensor was used).

The seven tightening levels can be used as a ground truth when designing learning methods. This makes this dataset useful for developing and testing clustering and classification methods for interpreting acoustic emission data.

The ORION-AE data are raw time-series. To be used in a VBGMM, we need a preprocessing stage. We used a similar preprocessing to [16] with a first step consisting in detecting the transients in the data stream and followed by a step of feature extraction. Conversely to [16], the Principal Components Analysis (PCA) was not used. Thirteen features were kept, and all combinations of four features were considered (four were used to decrease the amount of time of tests since all combinations were considered). The VBGMM was applied for each combination while also varying the amount of prior.

3.2 The priors

The priors (pl) were generated as proposed by Côme et al. [4, Section 5.2.1] using the true labels. For each training sample i, a number p_i was drawn from

a specific Beta distribution with expected value equal to $\rho \in \{0, 0.3, 0.6, 0.9, 1\}$ and variance 0.1, *used to define the doubt expressed by a hypothetical expert on the class of that sample*. With probability p_i, the label of sample i was changed (to any other class with equal probabilities). Therefore ρ controls the amount of prior introduced: When $\rho = 0$, all the labels are used as priors corresponding to supervised learning, whereas $\rho = 1$ corresponds to the unsupervised learning situation[1].

3.3 Sorting the partitions

The partitions obtained for all combinations for a given ρ were then sorted according to a criterion proposed in [14]. The criterion works as follows. For each partition, the onset time (first occurrence) of each cluster was determined. Then, each cluster was re-labelled according to their order of occurrence: the first cluster to occur was labelled "1", the second cluster labelled "2", and so on. This co-association allows the fusion of partitions since all clusters with the same label are assumed to correspond to the same source [13].

After re-labelling, each partition was ranked by :

$$\mathcal{C}(\mathcal{S}) = \sum_{k=1}^{\kappa-1} \Delta_{\text{onset}}(k, k+1) \log \Delta_{\text{onset}}(k, k+1) \tag{6a}$$

$$\Delta_{\text{onset}}(k, k+1) = t_{\text{onset,k+1}} - t_{\text{onset,k}} \tag{6b}$$

where κ is the number of cluster, $\mathcal{S}$ is a subset of features used to compute the partition, $t_{\text{onset},\kappa+1}$ is equal to the timestamp of the last AE signal. This criterion assumes that the onsets of all clusters in a given partition should be spread onto the time (or load) axis as uniformly as possible.

Once the partitions have been sorted according to this criterion, the best partition is taken for evaluating the quality of the clusters. In order to evaluate the performances, we used the Adjusted Rand Index [24], which provides a value between 0 and 1, where a "1" is obtained for perfect correspondence between the clusters estimated and the true labels.

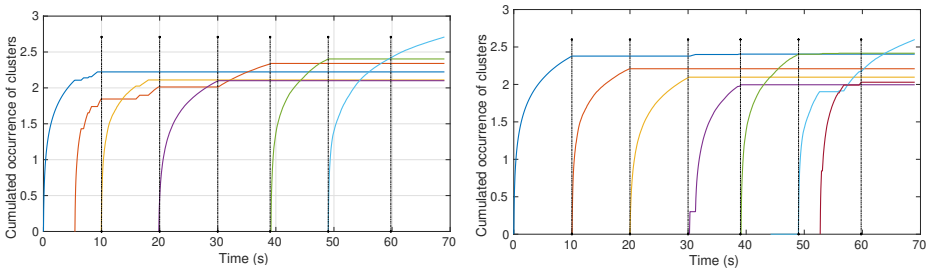
3.4 Results

Several tests were performed, using various ρ, considering uncertain and noisy priors, on the five datasets available in ORION-AE, with a comparison to clustering algorithms used in the literature. In this communication, results are only shown for the first dataset.

Figure 2 illustrates the results for the first dataset and for the second sensor. The priors were considered uncertain. Each curve represents the *decadic logarithm* of the cumulated number of acoustic emission transients in each clusters. For example, consider the left-hand side of the top-left figure (Fig. 2a, corresponding to the unsupervised case so with $\rho = 1$). The blue curve, for example,

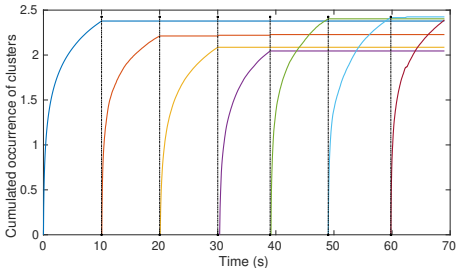
[1] The code can be found on T. Denœux's homepage in software/E2M/add_noise.m

is the cumulated number of the acoustic emission transients assigned to the first cluster. We can see that the curve reaches a plateau around 10 seconds, knowing that interval $[0, 10]$ seconds corresponds to the first level of tightening (the second interval, $[10, 20]$ is the second level, and so on). Therefore, this first cluster is relevant and can be assigned to the first tightening level because the blue curve evolves only in the first period. The yellow, purple, green and light blue clusters correspond to tightening levels 2, 3, 5 and 6 respectively. We can see that cluster 6 starting at 60 seconds does not stop increasing in $[60, 70]$ (level 7), which means that cluster 6 gathers data from both intervals. We can also see that the red cluster starts within $[0, 10]$ which means that the first level of tightening is split in two clusters. This red cluster stops increasing until 30 s where it grows again. This cluster is certainly related to the fourth level (the vertical axis is in logarithmic scale) and shares common features with the first one (according to the clustering method).

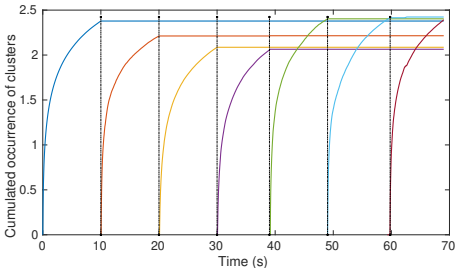


(a) $\rho = 1$ (without prior)

(b) $\rho = 0.9$



(c) $\rho = 0.6$



(d) $\rho = 0.3$

Fig. 2: Cumulated number of transient signals assigned to each cluster for $\rho = \{1,\ 0.9,\ 0.6,\ 0.3\}$.

The second figure (top right, Fig. 2b) corresponds to a small amount of prior, $\rho = 0.9$ (all labels are true but with large uncertainty). We can see that the red cluster is now at the good location, while a new cluster was positioned between 50 and 60 seconds. Therefore, there is a mixing between the two last levels. Then from $\rho = 0.6$ (Fig. 2c) downwards, the clusters correspond quite precisely to the tightening levels, with no difference between $\rho = 0.6$ and $\rho = 0.3$.

These results show that the cumulated plots of clusters bring two main information: The starting points (called onsets) of the accumulation, and the steady phase. When an onset is well located, it means that the clustering is able to assign the first transients of a given tightening level to the correct cluster. Concerning the steady phase, when it starts at the right location, it means that we are able to locate when a cluster stops increasing, therefore when a possible damage or functioning condition stops occurring. The height of the steady phase depends on how active was a damage and this can be useful for monitoring.

Figure 3 illustrates the evolution of the ARI for different values of ρ and considering all combinations of features (715). The ARI values were sorted by descending order. Circled-markers represent the 10 best ARI values corresponding to the 10 best partitions (estimated by the VBGMM-SOFT method and selected by the criterion proposed in [14]). We can see that the markers are located on the left-hand side of these curves corresponding to quite high ARI values.

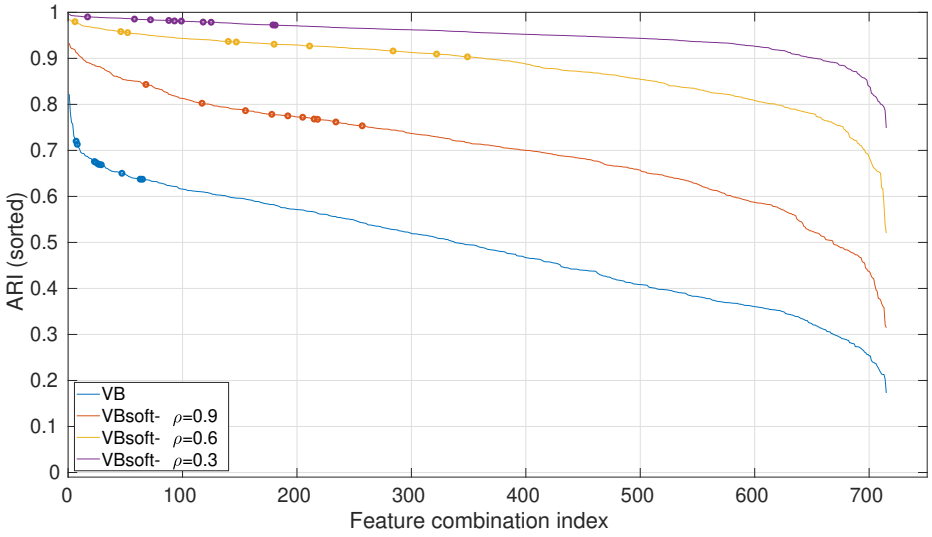


Fig. 3: Adjusted Rand Index (ARI) sorted for all partitions. The ten best partitions obtained by a VBGMM with different amounts of priors are superimposed with markers.

4 Conclusion

The first results obtained with a VBGMM on acoustic emission data are encouraging. We propose to add some prior which results in a small modification of the original algorithm.

With a small amount of prior the performance are greatly improved. The clusters obtained are interpreted by means of cumulated plots. For the application targeted, these plots underline two important pieces of information: the onsets of clusters and their steady phase. As a future work, we plan to exploit both of them for Structural Health Monitoring since it informs about the damage process taking place within the material.

The way to generate the prior remains a key problem. One idea followed by the acoustic emission community consists in getting prior knowledge from numerical simulations. However, in addition to the computational burden involved in such simulations, another difficulty holds in the fact that these simulations also require knowledge about the material properties which evolve during a test.

Acknowledgement

This work was supported by the EIPHI Graduate school (contract "ANR-17-EURE-0002").

References

1. Beal, M.J.: Variational Algorithms for Approximate Bayesian Inference. Ph.D. thesis, Gatsby Computational Neuroscience Unit, University College London (2003)
2. Bishop, C.: Pattern Recognition and Machine Learning. Springer (2006)
3. Chai, M., Zhang, J., Zhang, Z., Duan, Q., Cheng, G.: Acoustic emission studies for characterization of fatigue crack growth in 316ln stainless steel and welds. Applied Acoustics **126**, 101–113 (2017)
4. Côme, E., Oukhellou, L., Denoeux, T., Aknin, P.: Learning from partially supervised data using mixture models and belief functions. Pattern recognition **42**(3), 334–348 (2009)
5. Constantinou, A.C., Fenton, N., Neil, M.: Integrating expert knowledge with data in bayesian networks: Preserving data-driven expectations when the expert variables remain unobserved. Expert systems with applications **56**, 197–208 (2016)
6. Cour, T., Sapp, B., Taskar, B.: Learning from partial labels. The Journal of Machine Learning Research **12**, 1501–1536 (2011)
7. Denoeux, T.: Maximum likelihood estimation from uncertain data in the belief function framework. Knowledge and Data Engineering, IEEE Trans. on **25**(1), 119–130 (2013)
8. Denœux, T.: Maximum likelihood estimation from fuzzy data using the EM algorithm. Fuzzy sets and systems **183**(1), 72–91 (2011)
9. Giannì, C., Balsi, M., Esposito, S., Ciampa, F.: Low-power global navigation satellite system-enabled wireless sensor network for acoustic emission localisation in aerospace components. Structural Control and Health Monitoring **27**(6), e2525 (2020)
10. Grigg, S., Pullin, R., Pearson, M., Jenman, D., Cooper, R., Parkins, A., Featherston, C.A.: Development of a low-power wireless acoustic emission sensor node for aerospace applications. Structural Control and Health Monitoring **28**(4), e2701 (2021)

11. Omkar, S.N., Suresh, S., Raghavendra, T.R., Mani, V.: Acoustic emission signal classification using fuzzy c-means clustering. In: Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP '02. vol. 4, pp. 1827–1831 vol.4 (2002). https://doi.org/10.1109/ICONIP.2002.1198989
12. Quost, B., Denœux, T.: Learning from data with uncertain labels by boosting credal classifiers. In: Proceedings of the 1st ACM SIGKDD Workshop on Knowledge Discovery From Uncertain Data. pp. 38–47 (2009)
13. Ramasso, E., Placet, V., Boubakar, M.: Unsupervised consensus clustering of acoustic emission time-series for robust damage sequence estimation in composites. IEEE Trans. on Instr. and Meas. **64**(12), 3297–3307 (2015)
14. Ramasso, E., Butaud, P., Jeannin, T., Sarasini, F., Placet, V., Godin, N., Tirillò, J., Gabrion, X.: Learning the representation of raw acoustic emission signals by direct generative modelling and its use in chronology-based clusters identification. Engineering Applications of Artificial Intelligence **90**, 103478 (2020)
15. Ramasso, E., Denoeux, T.: Making use of partial knowledge about hidden states in hmms: an approach based on belief functions. IEEE Trans. on Fuzzy Systems **22**(2), 395–405 (2014)
16. Ramasso, E., Denoeux, T., Chevallier, G.: Clustering acoustic emission data streams with sequentially appearing clusters using mixture models. arXiv preprint arXiv:2108.11211 (2021)
17. Ramasso, E., Verdin, B., Chevallier, G.: Monitoring a bolted vibrating structure using multiple acoustic emission sensors: A benchmark. Data **7**(3), 31 (2022)
18. Sawan, H.A., Walter, M.E., Marquette, B.: Unsupervised learning for classification of acoustic emission events from tensile and bending experiments with open-hole carbon fiber composite samples. Composites Science and Technology **107**, 89–97 (2015)
19. Tonelli, D., Rossi, F., Luchetta, M., Caspani, V., Zonta, D., Migliorino, P., Selleri, A., Valeri, E., Marchiondelli, A., Ascari, G.: Effectiveness of acoustic emission monitoring for in-service prestressed concrete bridges. In: Huang, H., Zonta, D., Su, Z. (eds.) Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2021. vol. 11591, pp. 178 – 192. International Society for Optics and Photonics, SPIE (2021). https://doi.org/10.1117/12.2585527, https://doi.org/10.1117/12.2585527
20. Vannoorenberghe, P.: Finite mixture models estimation with a credal em algorithm. Traitement du Signal **24**(2), 103–113 (2007)
21. Vannoorenberghe, P., Smets, P.: Partially supervised learning by a credal EM approach. In: Godo, L. (ed.) Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Lecture Notes in Computer Science, vol. 3571, pp. 956–967. Springer Berlin Heidelberg (2005)
22. Vendramin, L., Campello, R., Hruschka, E.: Relative clustering validity criteria: A comparative overview. Statistical Analysis and Data Mining **3**(4), 209–235 (2010)
23. Vendramin, L., Jaskowiak, P.A., Campello, R.J.: On the combination of relative clustering validity criteria. In: Proceedings of the 25th Int. Conf. on Scientific and Statistical Database Management. p. 4. ACM (2013)
24. Vinh, N., Epps, J., Bailey, J.: Information theoretic measures for clusterings comparison: Is a correction for chance necessary? In: Proc. of the 26th Annual Int. Conference on Machine Learning. pp. 1073–1080. ACM, New York, NY, USA (2009)