

Soft-ECM : une extension de l’algorithme Evidential C-Means pour des données complexes

Armel Soubeiga*, Thomas Guyet**
Violaine Antoine*

*Université Clermont Auvergne, Clermont Auvergne INP, UMR 6158 CNRS, LIMOS,
F-63000 Clermont-Ferrand, France
armel.soubeiga@uca.fr

** Inria, AISTroSight, Lyon, France
thomas.guyet@inria.fr

Résumé. Dans cet article, nous reformulons le problème de l’*Evidential C-Means* (ECM) pour le clustering de données complexes. Nous proposons un nouvel algorithme, Soft-ECM, qui positionne de manière cohérente les centres des clusters imprécis en se basant sur une semi-métrique. Nos expérimentations montrent que Soft-ECM obtient des résultats comparables aux approches usuelles de clustering floue sur des données dans un espace euclidien et nous montrons sur des données catégorielles et des séries temporelles l’intérêt de combiner le clustering floue et l’utilisation de semi-métriques.

1 Introduction

Dans de nombreuses applications réelles, les caractéristiques des objets peuvent être peu discriminantes, dans ces cas, le clustering rigide peut entraîner une précision insuffisante. Pour résoudre ce problème et capturer le degré d’ambiguïté, des variantes de clustering souple ont été proposées, notamment Evidential C-Means (Masson et Denœux, 2008). Ces variantes permettent d’exprimer l’incertitude et/ou l’imprécision dans la partition. Cependant, ECM a été conçu pour des données tabulaires numériques. Il ne peut pas être appliqué à des données complexes telles que des séries temporelles, pour lesquelles des méthodes de clustering dur ont déjà été adaptées (Petitjean et al., 2014). ECM repose sur l’utilisation de la distance euclidienne comme métrique de référence, ce qui limite sa capacité à traiter des données dans des espaces non-euclidiens. Ceci a motivé la proposition de variantes cherchant à adapter ECM à différents types de données. Parmi ces variantes, ECMdd (Zhou et al., 2015) utilise des mesures de similarité relationnelles, tandis que CECM (Antoine et al., 2012) introduit des contraintes de clustering pour mieux orienter la partition des données et emploie une distance de Mahalanobis spécifique pour chaque clusters. D’autres variantes, comme CatECM (Djiberou Mahamadou et al., 2019), sont spécifiquement conçues pour des données catégorielles.

Dans cet article, nous proposons Soft-ECM, un nouvel algorithme qui introduit plusieurs améliorations. D’abord, nous reformulons le problème en un clustering multi-classes contraint, où tous les clusters, qu’ils soient singletons ou méta-clusters (clusters imprécis), sont optimisés individuellement. Contrairement à l’approche classique d’ECM, les méta-clusters sont ici

considérés comme des clusters à part entière, avec leur propre optimisation, ce qui permet une meilleure flexibilité dans la gestion des incertitudes. En contrepartie, nous introduisons une nouvelle contrainte pour assurer que les barycentres des méta-clusters soient positionnés de manière cohérente par rapport à ceux des clusters singletons. De la sorte, l'algorithme Soft-ECM est plus adaptable, et permet l'utilisation de semi-métriques. Cela permet de traiter efficacement des données complexes, comme celles qui présentent des structures cachées ou qui ne s'ajustent pas naturellement à la distance euclidienne.

2 Rappels sur Evidential C-Means (ECM)

ECM est un algorithme qui s'appuie sur les notions d'incertitude et d'imprécision définies dans la théorie de Dempster-Shafer (Shafer, 1976). Considérant ω une variable qui prend des valeurs dans un ensemble fini $\Omega = \{\omega_1, \dots, \omega_c\}$, une connaissance partielle de la valeur réelle de ω peut être représentée par une fonction de masse m , définie de l'ensemble 2^Ω vers $[0, 1]$ et satisfaisant : $\sum_{A \subseteq \Omega} m(A) = 1$. Lorsque $m(A) = 1$, nous avons une croyance certaine. Lorsque $m(A) > 0$ et $|A| > 1$, nous avons une croyance imprécise.

Le clustering d'ECM (Masson et Denœux, 2008) considère $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ un ensemble de n individus décrits dans $\mathbb{R}^p$ à classer dans un ensemble de c clusters (« purs »), noté $\Omega = \{\omega_1, \dots, \omega_c\}$. Le résultat d'ECM est alors l'assignation de chaque exemple $\mathbf{x}_i \in \mathbf{X}$ à un sous-ensemble $A \subseteq \Omega$ par une masse $m_i(A) \in \mathbb{R}_+$. Lorsque A est de cardinalité strictement supérieure à un ($|A| > 1$), un exemple est classé de manière incertaine aux clusters de cet ensemble. A est alors désigné comme un *méta-cluster*.

L'algorithme ECM consiste à minimiser l'inertie intra-classe suivante :

$$J_{ECM}(\mathbf{M}, \mathbf{V}) = \sum_{i=1}^n \sum_{A \subseteq \Omega \setminus \emptyset} |A|^\alpha m_i(A)^\beta d^2(\mathbf{x}_i, \bar{\mathbf{v}}_A) + \sum_{i=1}^n \delta^2 m_i(\emptyset)^\beta, \quad (1)$$

sous les contraintes $\sum_{A \subseteq \Omega} m_i(A) = 1 \quad \forall i = 1, \dots, n$

avec $\mathbf{M} = \{m_i(A) \mid i \in [n], A \subseteq \Omega\}$ la partition nommée partition crédale, $\mathbf{V} = \{\mathbf{v}_{\omega_j}\}_{j \in [c]}$ les centroïdes des clusters « purs », $\{\bar{\mathbf{v}}_A\}_{A \subseteq \Omega}$ les centroïdes des clusters et méta-clusters (définis ci-dessous); et d est la distance euclidienne.

Il est important de noter que les centroïdes des méta-clusters ne font pas partie de $\mathbf{V}$, mais ils sont calculés comme barycentre des centres associés aux clusters de chaque sous-ensemble $A \subseteq \Omega \setminus \emptyset$:

$$\bar{\mathbf{v}}_A = \frac{1}{|A|} \sum_{\omega \in A} \mathbf{v}_\omega. \quad (2)$$

On constate alors que ECM exploite les propriétés de la distance euclidienne aussi bien pour la définition de la fonction à optimiser (eq. 1) ou la définition des barycentres (eq. 2). Ceci rend la méthode inadaptée pour des données complexes telles que des données catégorielles, non-numériques, ou des séries temporelles, pour lesquels on souhaiterait, par exemple, utiliser une semi-métrique de type DTW (Cuturi et Blondel, 2017).

3 L'approche Evidential C-Means souple (Soft-ECM)

Dans cette partie, $\mathbf{X} = \{x_1, \dots, x_n\}$ désigne une collection d'objets dans un espace $\mathbb{X}$ (e.g., des séries temporelles) et d désigne ici une semi-métrique entre ces objets. De manière classique (Petitjean et al., 2014), on peut alors définir le barycentre d'un ensemble d'objets comme un élément de $\mathbb{X}$ qui minimise l'inertie. Pour un sous-ensemble $E \subseteq \mathbf{X}$, son barycentre, $\mathbf{v}_E$, peut être défini par :

$$\mathbf{v}_E = \operatorname{argmin}_{x \in \mathbb{X}} \sum_{e \in E} d(e, x). \quad (3)$$

Notre approche relaxe la contrainte dure de définition des centres de meta-cluster. Nous définissons le problème Soft-ECM par le problème de minimisation sous contrainte ci-dessous :

$$\begin{aligned} \min_{\mathbf{M}, \mathbf{V}} J^{\alpha, \beta, \delta}(\mathbf{M}, \mathbf{V}) &= \sum_{i=1}^n \sum_{A \subseteq \Omega \setminus \emptyset} |A|^\alpha m_i(A)^\beta d(x_i, \mathbf{v}_A) + \sum_{i=1}^n \delta^2 m_i(\emptyset)^\beta \\ \text{sous les contraintes} \quad &\sum_{A \subseteq \Omega} m_i(A) = 1, \quad \forall i \in \{1, n\}, \\ &\mathbf{v}_A = \operatorname{argmin}_{x \in \mathbb{X}} \sum_{\omega_k \in A} d(\mathbf{v}_{\{\omega_k\}}, x), \quad \forall A \subseteq \Omega. \end{aligned} \quad (4)$$

Dans cette formulation, $\mathbf{M} = (m_i(A))$ représente toujours les fonctions de masse sur les éléments focaux $A \subseteq \Omega$, mais $\mathbf{V} = \{\mathbf{v}_A\}$ correspond aux barycentres des clusters et des méta-clusters. Contrairement à la formulation d'ECM (eq. 1) qui cherchait à optimiser uniquement les positions des clusters, définissant par l'équation 2 les positions des méta-clusters, on optimise ici les deux à la fois. En contrepartie de cette relaxation, la seconde contrainte impose au centroïde de chaque méta-cluster A de se positionner comme le barycentre des centroïdes des clusters ($\mathbf{v}_{\omega_k}$ où $\omega_k \in A$). On retrouve ici l'expression de l'équation 3 qui définit le barycentre.

Dans le cas particulier où $\mathbb{X} = \mathbb{R}^p$ et d est une distance euclidienne, les problèmes des équations 1 et 4 sont équivalents.

4 Résolution de Soft-ECM

Le problème Soft-ECM est un problème d'optimisation bilevel (Dempe et Zemkoho, 2020), dont nous proposons une heuristique de résolution qui relaxe la contrainte dure d'identification des barycentres de méta-clusters en une contrainte numérique additionnelle dans la fonction globale à minimiser. Le problème de minimisation de Soft-ECM relaxé devient alors :

$$\begin{aligned} \min_{\mathbf{M}, \mathbf{V}} \sum_{i=1}^n \sum_{A \subseteq \Omega \setminus \emptyset} |A|^\alpha m_i(A)^\beta d(x_i, \mathbf{v}_A) &+ \sum_{i=1}^n \delta^2 m_i(\emptyset)^\beta + \lambda \sum_{A \subseteq \Omega \setminus \emptyset} \sum_{\omega_k \in A} d(\mathbf{v}_{\{\omega_k\}}, \mathbf{v}_A) \\ \text{sous la contrainte} \quad &\sum_{A \subseteq \Omega} m_i(A) = 1, \quad \forall i \in \{1, n\}. \end{aligned} \quad (5)$$

Le premier et le second termes sont similaires à ceux de ECM, tandis que le troisième terme impose la contrainte supplémentaire sur la cohérence entre les barycentres des clusters

$\{\omega_k\}$ et ceux des méta-clusters. Le nouvel hyper-paramètre λ ajuste l'importance donnée à cette contrainte.

Le problème d'optimisation relaxé de Soft-ECM peut être résolu à l'aide d'un schéma d'optimisation alternée, en alternant (jusqu'à convergence ou un nombre d'itérations fixé) les deux étapes suivantes :

Optimisation par rapport à M pour $\mathcal{V}$ fixé. La partition crédale M , qui minimise la fonction objectif est calculée de manière explicite par l'expression ci-dessous :

$$m_i(A) = \frac{|A|^{-\frac{\alpha}{\beta-1}} d(x_i, \mathbf{v}_A)^{-\frac{1}{\beta-1}}}{\sum_{B \neq \emptyset} |B|^{-\frac{\alpha}{\beta-1}} d(x_i, \mathbf{v}_B)^{-\frac{1}{\beta-1}} + \delta^{-\frac{2}{\beta-1}}}, \quad (6)$$

$$m_i(\emptyset) = 1 - \sum_{A \neq \emptyset} m_i(A).$$

Cette solution a été proposée par (Masson et Denœux, 2008) pour ECM et reste valide pour Soft-ECM. En effet, le terme ajouté à la fonction objectif ne dépendant pas de M , l'optimisation par rapport à M du problème de l'eq. 5 est équivalent à celle du problème de l'eq. 1.

Optimisation par rapport à $\mathcal{V}$ pour M fixé. L'optimisation de l'eq. 5 lorsque M est fixé est équivalente au problème d'optimisation non contrainte suivant :

$$\underset{\mathcal{V}}{\operatorname{argmin}} \sum_{A \subseteq \Omega \setminus \emptyset} \left(\sum_{i=1}^n |A|^\alpha m_i(A)^\beta d(x_i, \mathbf{v}_A) + \lambda \sum_{\omega_k \in A} d(\mathbf{v}_{\{\omega_k\}}, \mathbf{v}_A) \right) \quad (7)$$

Cette fois, nous adoptons une approche différente de celle de (Masson et Denœux, 2008) et utilisons une approche numérique (descente de gradient) pour trouver une solution approchée. Les outils d'optimisation basés sur l'*automatic differentiation* résolvent efficacement ce problème.

5 Expérimentations

Nous commençons par illustrer le comportement de Soft-ECM par rapport à ECM à l'aide d'un ensemble de données classique appelé *Diamond* (Masson et Denœux, 2008). Il est composé de douze objets (cf. Figure 1a), dont les objets 1 à 11 sont normaux, tandis que l'objet 12 est une valeur aberrante. Soft-ECM et ECM ont été exécutés avec les mêmes paramètres : $\alpha = \frac{1}{6}$, $\beta = 2$, $\delta = 11$ et, pour Soft-ECM, $\lambda = 1.5$. Une partition en $c = 2$ clusters a été imposée. Les Figures 1b et 1c présentent les masses des 2^c clusters (éléments focaux) obtenues dans le processus d'optimisation de clustering – $m(\omega_1)$, $m(\omega_2)$, $m(\Omega)$ et $m(\emptyset)$ – pour ECM et Soft-ECM.

Les objets 1 à 5 et 8 à 11 sont classés dans deux clusters distincts par ECM et Soft-ECM, tandis que l'objet 12 est identifié comme aberrant. Pour les objets 6 et 7, ECM concentre la majeure partie de la masse sur l'ignorance Ω , car ils sont proches du centre de Ω , défini comme la moyenne des centres de ω_1 et ω_2 (cf. Figure 1b). En revanche, Soft-ECM optimise mieux ce centre, avec une masse de croyance majoritairement assignée à ω_2 pour l'objet 7 (cf. Figure 1c).

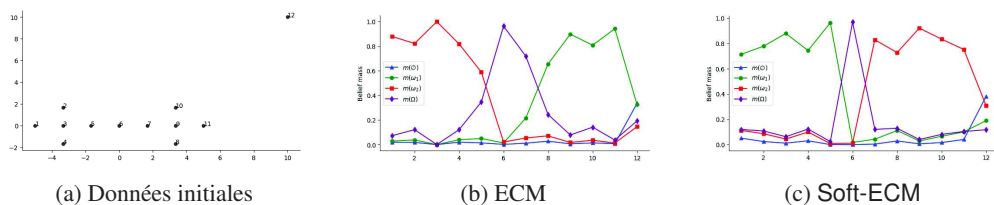


FIG. 1 – Illustration du regroupement de l'ensemble de données sur des données jouets (Masson et Denœux, 2008). (a) illustration des données. (b) et (c) distributions des masses pour chaque objet pour ECM et Soft-ECM. Plus la masse est élevée et plus un exemple est estimé appartenir à un cluster ou à un meta-cluster.

Type	Données	n	p	c	Longueur	Métrie(s)
Mixtes	Abalone	4177	8	3	-	Euclidienne
Mixtes	Ecoli	336	7	8	-	Euclidienne
Catégoriel	Breast Cancer (BC)	286	9	2	-	Hamming
Série T.	StandWalkJump (SWJ)	12	4	3	2500	Euclidienne, Soft-DTW
Série T.	BasicMotions (BM)	40	6	4	100	Euclidienne, Soft-DTW

TAB. 1 – Caractéristiques des ensembles de données expérimentaux, n nombre d'objets, p nombre de dimensions et c le nombre de classes.

Nous évaluons maintenant les performances de l'algorithme Soft-ECM en le comparant à ECM et à ses diverses variantes issues de l'état de l'art, en utilisant des ensembles de données réelles aux formats variés. Nous avons mené les expériences sur cinq jeux de données publics¹ de différents types (mixtes, catégoriels et séries temporelles). Ces jeux sont décrits dans la Table 1.

L'algorithme Soft-ECM² a été implémenté en Python et s'appuie sur le framework PyTorch (Paszke et al., 2017) pour la résolution de l'éq. 7. Les algorithmes de référence utilisés pour comparer les performances de notre proposition sont l'algorithme ECM pour les données mixtes, l'algorithme Evidential C-Means pour les données catégorielles (CatECM) (Djiberrou Mahamadou et al., 2019)), et l'algorithme Evidential C-Medoids (ECMdd) (Zhou et al., 2015) pour les données temporelles. Nous avons également comparé nos résultats avec une approche de clustering dure, à savoir K-means et ses variantes K-modes et TimeSeriesKMeans (TSKmeans) (Fahad et al., 2014).

Chaque algorithme de clustering a été exécuté jusqu'à la convergence vers une valeur stationnaire avec $\epsilon = 10^{-3}$ avec les paramétrages suivants : $\alpha = 2.0$, $\delta = 10$, $\beta \in \{1.1, 1.2, \dots, 2.0\}$, $\lambda \in \{1.0, 2.0, \dots, 10.0\}$, et le nombre de sous-ensembles pour les méta-clusters a été limité à 2 pour des raisons computationnelles et d'interprétabilité des résultats. Pour chaque ensemble de données, le nombre de clusters correspond au nombre de clusters a priori. Concernant les partitions crédales, la transformation pignistique (Masson et Denœux, 2008) a été uti-

1. <https://archive.ics.uci.edu>, <https://www.timeseriesclassification.com/>

2. Code disponible : <https://gitlab.inria.fr/tguyet/ecm-ts>.

	ECM	ECMdd	CatECM	Soft-ECM	K-Means	K-Modes	TSKmeans
Adjusted Rand Index							
Abalone	0.58	0.57	-	0.57	0.56	-	-
Ecoli	0.88	0.84	-	0.87	0.87	-	-
BC	-	0.58	0.58	0.59	-	0.58	-
SWJ (Euclidean)	-	0.41	-	0.47	-	-	0.45
BM (Euclidean)	-	0.51	-	0.62	-	-	0.4
SWJ (SoftDTW)	-	0.35	-	0.47	-	-	0.47
BM (SoftDTW)	-	0.38	-	0.62	-	-	0.39
Accuracy							
Abalone	0.51	0.48	-	0.52	0.51	-	-
Ecoli	0.79	0.732	-	0.78	0.81	-	-
BC	-	0.70	0.70	0.72	-	0.70	-
SWJ (Euclidean)	-	0.5	-	0.5	-	-	0.5
BM (Euclidean)	-	0.4	-	0.45	-	-	0.4
SWJ (SoftDTW)	-	0.42	-	0.5	-	-	0.5
BM (SoftDTW)	-	0.37	-	0.45	-	-	0.37

TAB. 2 – Précision de clustering (Rand Index et Accuracy) des algorithmes sur les jeux de données réels. Les chiffres en gras indiquent les meilleurs résultats par jeu de données. Les “-” indiquent des méthodes inadaptées à un jeu de données.

lisée afin de les convertir en partitions dures. Les (semi-)métriques euclidienne, Soft-DTW et Hamming ont été utilisées respectivement pour le clustering des données mixtes, temporelles et catégorielles (voir Table 1). Deux mesures de performance standard sont calculées sur ces partitions : la précision (*accuracy*) et l’indice de Rand ajusté (ARI).

La Table 2 présente les résultats de précision obtenus pour chaque algorithme sur chaque jeu de données. Les précisions sont calculées par rapport aux catégories connues des exemples. Pour chaque algorithme, on a sélectionné les résultats du paramétrage ayant le coût minimal de la fonction objectif (voir analyse de sensibilité ci-dessous).

Ces résultats montrent tout d’abord que Soft-ECM a des performances comparables aux autres méthodes sur les données mixtes partitionnées avec une distance euclidienne. Soft-ECM arrive second derrière *ECM* pour le critère de mesure ARI (soit en moyenne 72.3% de bon regroupement pour Soft-ECM contre 72.9% pour *ECM*). Il se classe troisième en terme d’*accuracy* avec une moyenne de 64.8%, derrière *KMeans* (66.2%) et *ECM* (65.2%). Nous nous attendions à ces résultats dans la mesure où le problème de Soft-ECM est équivalent à celui d’*ECM* lorsqu’on utilise une distance euclidienne.

Le résultat intéressant est obtenu sur les données temporelles et catégorielles : Soft-ECM est plus précis, en ARI et en *accuracy*, y compris contre les algorithmes de clustering durs dédiés. En effet, on constate que Soft-ECM a des ARI supérieures à *K-Modes* et *TSKmeans* respectivement pour le partitionnement dur des données catégorielles et temporelles. Ces per-

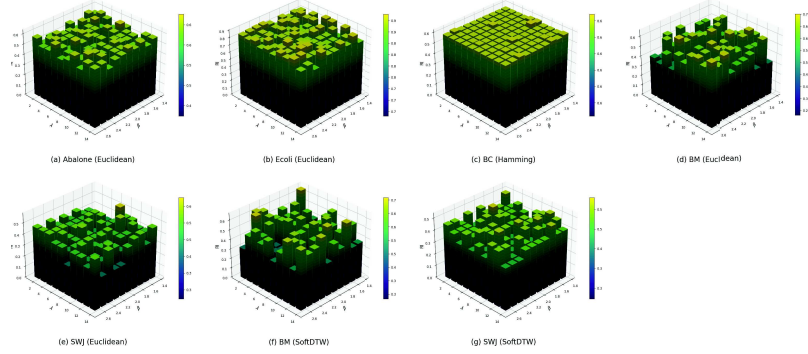


FIG. 2 – Sensibilité aux paramètres β et λ sur la performance de l'algorithme Soft-ECM (ARI).

formances ne sont pas uniquement dues à la prise en compte de l'incertain, puisque Soft-ECM est également meilleur que *CatECM* pour le clustering crédale des données catégorielles.

On conclut de cette expérimentation que l'algorithme proposé présente des performances comparables ou supérieures aux algorithmes de référence utilisés dans cette expérimentation comparative.

On s'intéresse finalement à la sensibilité des résultats de Soft-ECM au paramètre λ , paramètre additionnel par rapport à la méthode *ECM*. Les mesures d'ARI pour différentes combinaisons de valeurs de λ et β sont présentées dans la Figure 2.

On peut constater que les valeurs de λ et β ont un impact significatif sur les performances de Soft-ECM. Ces premiers résultats ne montrent pas de tendance évidente dans l'évolution de l'ARI. Le choix du paramètre dépend fortement de la nature des données et, à ce stade, le choix des paramètres doit se réaliser de manière manuelle.

6 Conclusion

Dans cet article, nous avons proposé une généralisation de l'algorithme Evidential C-Means, appelée Soft-ECM, afin d'apporter une solution de clustering flou adaptée à des données complexes, incertaines et non nécessairement définies dans un espace euclidien. Les résultats expérimentaux montrent que Soft-ECM est capable de produire des performances comparables, voire supérieures, à celles des algorithmes de référence. De plus, il se distingue par sa capacité à offrir une représentation plus fine du flou, un atout majeur pour l'analyse de données incertaines. Bien que cette extension introduise de nouveaux hyper-paramètres et qu'elle repose sur un schéma de résolution numérique, la flexibilité apportée par Soft-ECM dans l'utilisation de semi-métriques et l'amélioration du calcul des barycentres des méta-clusters le rendent particulièrement pertinent pour une large gamme de problèmes. En perspective, nous envisageons d'améliorer la relaxation du problème en utilisant une optimisation bi-niveau. De plus, une étude approfondie de la complexité temporelle ainsi qu'un élargissement des expérimentations sur d'autres jeux de données complexes constituent des pistes de développement pour évaluer davantage la robustesse et l'efficacité de l'algorithme.

Remerciements

A. Soubeiga et V. Antoine remercient l'Agence nationale de la recherche française pour son soutien dans le cadre du programme Investissements d'avenir (16-IDEX-0001 CAP 20-25).

Références

- Antoine, V., B. Quost, M.-H. Masson, et T. Denœux (2012). CECM : Constrained evidential C-Means algorithm. *Computational Statistics & Data Analysis* 56(4), 894–914.
- Cuturi, M. et M. Blondel (2017). Soft-DTW : a differentiable loss function for time-series. In *Proceedings of the International Conference on Machine Learning (ICML)*, Volume 70, pp. 894–903.
- Dempe, S. et A. Zemkoho (2020). Bilevel optimization. In *Optimization and its applications*, Volume 161. Springer.
- Djiberou Mahamadou, A. J., V. Antoine, G. J. Christie, et S. Moreno (2019). Evidential clustering for categorical data. In *Proceedings of the International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1–6.
- Fahad, A., N. Alshatri, Z. Tari, A. Alamri, I. Khalil, A. Y. Zomaya, S. Foufou, et A. Bouras (2014). A survey of clustering algorithms for big data : Taxonomy and empirical analysis. *IEEE Transactions on Emerging Topics in Computing* 2(03), 267–279.
- Masson, M.-H. et T. Denœux (2008). ECM : An evidential version of the fuzzy c-means algorithm. *Pattern Recognition* 41(4), 1384–1397.
- Paszke, A., S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, et A. Lerer (2017). Automatic differentiation in PyTorch. In *Proceedings of the Conference on Neural Information Processing Systems (NIPS)*.
- Petitjean, F., G. Forestier, G. I. Webb, A. E. Nicholson, Y. Chen, et E. Keogh (2014). Dynamic time warping averaging of time series allows faster and more accurate classification. In *Proceedings of the International conference on data mining (ICDM)*, pp. 470–479.
- Shafer, G. (1976). *A mathematical theory of evidence*, Volume 42. Princeton university press.
- Zhou, K., A. Martin, Q. Pan, et Z. Liu (2015). Evidential relational clustering using medoids. In *Proceedings of the International Conference on Information Fusion (Fusion)*, pp. 413–420.

Summary

In this paper, we reformulate the *Evidential C-Means* (ECM) problem for clustering complex row-data. We propose a new algorithm, Soft-ECM, which coherently positions the centers of imprecise clusters based on a semi-metric. Our experiments show that Soft-ECM obtains results comparable to conventional fuzzy clustering approaches on data in a euclidean space, and we demonstrate its benefits when combining fuzzy clustering with semi-metrics on time series data or categorical relational data.