

Extraction de profils de patients à partir de données de cardiomyopathie héréditaire

Février 2024

Résumé : Notre projet d'ouverture à la recherche aborde le clustering multi-vues qui sera défini plus tard. En collaboration avec l'hôpital de Bron qui nous fournit les données dans le cadre du projet CARDIA, encadré par Dr Khalid Benabdeslem, et co-encadré par le Dr Haytham Elghazel. Les données décrivent des patients sujet à la cardiomyopathie, une maladie du cœur, héréditaire.

Mots-clés : clustering, unsupervised, multi-view, cardia, cardiomyopathie

1. Présentation du sujet

Les Cardiomyopathies[W1] concernent notamment des millions de jeunes sportifs à travers le monde. A ce jour, plus d'un millier de mutations ont été identifiées dans de nombreux gènes, indiquant que diverses molécules et voies entraînent des cardiomyopathies hypertrophiques, dilatées, restrictives et arythmogènes.

Ces mutations sont souvent spécifiques à un patient, pertinentes pour une personne, ou spécifiques à un groupe, c'est-à-dire caractéristiques de populations de patients sous-spécifiques.

Cependant, l'hétérogénéité génétique et la variation allélique associées à la variation de l'expression clinique posent des défis importants aux cliniciens face à un individu. De plus, le phénotypage faible limite la puissance de l'analyse génétique.

Les données récoltées constituent en effet, un gisement considérable de signaux cardiaques, (électrocardiogramme(ECG)) et de variantes génétiques, en guise d'un phénotypage clinique pertinent sur la maladie, obtenu grâce à un profiling clinico-biologique précis pour les professionnels de la santé. Il est donc primordial pour ces derniers de disposer de la connaissance permettant d'une part, de diagnostiquer au mieux la cardiomyopathie héréditaire et d'autre part, de servir comme outil d'aide à la décision thérapeutique.

Notre objectif sur un an de Janvier 2024 à Décembre 2024(cf. figure 7), est d'établir des groupes de patients avec un contexte multi-vues, c'est-à-dire de mettre en alliance, simultanément, les caractéristiques cliniques et les caractéristiques biologiques pour former des clusters de patients homogènes, pouvant afficher une image complète exprimant des symptômes similaires.

Pour répondre à ce problème, l'approche réalisée est un clustering non supervisé multi-vues. Non supervisé signifie que notre modèle n'a pas de connaissances sur le diagnostic du patient c'est-à-dire qu'on ne sait pas si il est malade ou non.

L'approche multi-vues signifie que le patient est décrit dans plusieurs contextes à savoir différents types d'examens médicales:

- Clinique
- Génétique
- Électrocardiogramme (ECG)

Dans ce rapport nous allons établir un état de l'art concernant les méthodes connues pour répondre à ce type de problème. Nous nous appuyons sur plusieurs papiers de recherches récents. Nous avons pour objectif d'ici décembre 2024 de faire une étude applicative (en Python) des approches investiguées avec des tests statistiques à l'appui. On utilisera aussi un jeu de données de type similaire trouvé sur internet afin d'enrichir notre comparaison des modèles.

2. Description des données

L'hôpital de Cardiologie de Bron nous a déjà fourni un échantillon des données. Après une analyse de cet échantillon on remarque plusieurs données manquantes comme ce qui est courant dans la science des données. Les données manquantes devront être traitées afin qu'elles n'apportent pas d'incohérence.

2.1. Clinique

Les données cliniques comprennent des informations du patient telles que les dates de ses précédents examens, ses antécédents familiaux. C'est la base qui nous permettra de faire le lien avec l'identifiant dans vues.

2.2. Génétique

L'échantillon de la donnée génétique décrit certaines informations sur le génome d'un patient :

id	Category	gene	overlapKnown	type	codingConsequence	refGenome	chromosome	genome_position	depth	var_percent
235A		TTN		INDEL	frameshift	GRCh37/hg19		2	179439702	145439.13
623B		AKAP9		INDEL	inframe_3	GRCh37/hg19		7	91652178	138243.92
1007B		CRYAB		SNP	splice_donor_+4	GRCh37/hg19		11	111781047	142949.9

figure 1 : échantillon de la liste des variants génétiques

Le nombre de variables est grand ce qui implique une grande dimension, c'est une complexité supplémentaire qui nécessitera une sélection des variables pertinentes, et le choix d'un algorithme adapté à une grande dimension.

2.3. ECG

L'electrocardiogramme est l'une des méthodes utilisée par les experts en cardiologie la plus répandue pour déterminer la cardiomyopathie chez le patient. Notre échantillon nous présente un fichier XML du logiciel Muse.

La classification d'un ECG nécessite une grosse préparation de la donnée, en effet on doit nettoyer les effets de distorsions et le bruit induit par la respiration du patient, les contractions des muscles, etc...
Ensuite La méthode couramment utilisée pour la classification d'ECG est de segmenter l'ECG de façon périodique et d'annoter les courbes identifiables (P, QRS, T, etc...).

Nous avons étudié un récent papier de recherche réalisant l'état de l'art autour de la classification non-supervisée pour l'analyse des ECG. Il cite plusieurs algorithmes basiques montrant une forte performance (Accuracy), parmi un clustering K-means + SVD, mais aussi des algorithmes plus récents en deep-learning qui présentent des performances plus restreintes. Ces affirmations seront à démontrer dans notre future comparaison applicative.

Dans un second papier de recherche, nous avons étudié leur méthode comprenant le clustering floue, une autre méthode décrite dans l'état de l'art suivant.

3. État de l'art

Le Clustering non supervisée est un concept de machine learning visant à créer des groupes d'individus en apprenant sur la donnée. Le premier algorithme présenté permet de bien comprendre le concept.

3.1. Clustering

Pour faciliter la compréhension des algorithmes, nous allons établir quelques définitions afin de rester au plus proche du vocabulaire technique :

Classes / Clusters	un ensemble d'individu représentant un groupes
Outliers	Donnée aberrante

3.1.1. K-Means

L'algorithme le plus utilisé est K-Means [IEAAJ23]. Il vise à minimiser l'inertie intra-classe c'est-à-dire minimiser la distances des individus au sein d'un groupe, mais aussi maximiser l'inertie inter-classe de ces groupes. Il est demandé un nombre de groupes à l'initialisation de l'algorithme.

Nous allons décrire son fonctionnement qui est très simple. Il commence par initialiser aléatoirement un ensemble de centroïdes, qui représentent les centres des groupes. Ensuite K-Means alterne entre deux étapes, jusqu'à arriver à une configuration de cluster qui est stable, c'est-à-dire qu'elle ne change pas entre deux itérations.

- **Étape d'affectation** : Les points de données sont assignés au cluster dont le centroïde est le plus proche, généralement en utilisant la distance euclidienne.
- **Étape de mise à jour des centroïdes** : Les centroïdes sont recalculés en prenant la moyenne des points de données attribués à chaque cluster.

Nous allons présenter dans l'exemple ci-dessous comment l'algorithme fonctionne avec un exemple de 3 classes.

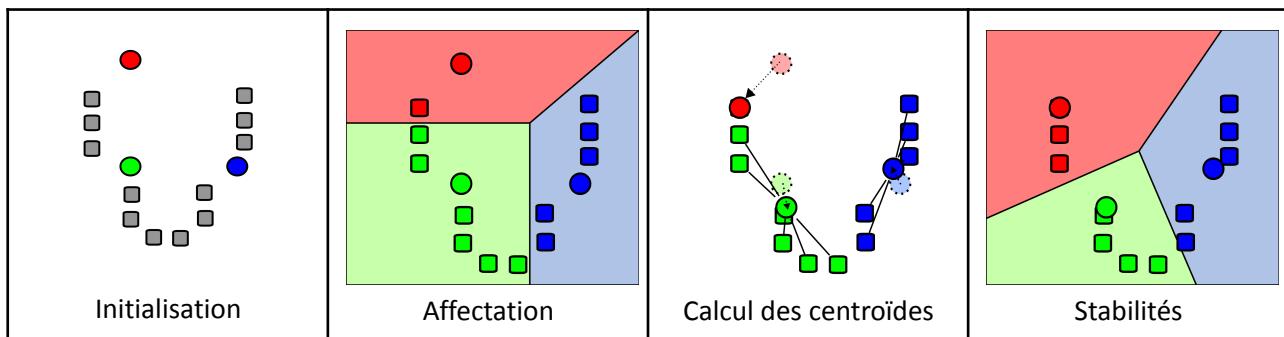


Figure 2 : Etapes du Kmeans (from wikipedia)

Avantages de la méthode K-means :

- **Simplicité et efficacité** : La méthode K-Means est simple à comprendre et à appliquer, de plus cette méthode est rapide, en effet sa complexité temporelle est linéaire.
- **Garantie de convergence** : K-means garantit la convergence vers un minimum local de la fonction objectif. Cela signifie que même si la solution trouvée n'est pas nécessairement optimale globalement, elle est au moins stable.

Inconvénients de la méthode K-means :

- **Sensibilité à l'initialisation aléatoire des centroïdes** : Les résultats de la méthode K-means peuvent varier en fonction de l'initialisation des centroïdes. Une mauvaise initialisation peut conduire à une convergence vers un minimum local sous-optimal.
- **Sensibilité aux outliers** : K-means est très sensibles aux outliers (donnée aberrante), et la présence de ces outliers dans les données peuvent significativement réduire sa performance.
- **Nombre de clusters à préciser** : un autre inconvénient est qu'il faut préciser le nombre de clusters à l'algorithme, ce que l'on ne connaît pas toujours.
- **Possibilité de vider un groupe** : Il est possible de vider les classes avec une mauvaise initialisation.

Pour pallier ces faiblesses, il y a plusieurs variantes de K-Means présenté en annexe.(cf. figure 8)

3.1.2. CAH

Le clustering ascendant hiérarchique (CAH) est une méthode de clustering populaire qui ne nécessite pas de connaître le nombre de cluster à l'avance.[W2] La CAH consiste à attribuer à chaque point de données son propre cluster. Ensuite, il fusionne récursivement les clusters les plus similaires jusqu'à ce qu'un seul cluster global soit formé. Cette fusion est basée sur une mesure de similarité ou de dissimilarité entre les clusters, souvent calculée à l'aide de mesures telles que la distance euclidienne, la distance de Manhattan ou la corrélation.

Méthodes de fusions :

- **Agglomérative** : L'approche agglomérative est ascendante où chaque observation commence dans son propre cluster, et les paires de clusters sont fusionnées à mesure que l'on remonte dans la hiérarchie. Cette méthode est utilisée pour la segmentation d'image, l'analyse des réseaux sociaux, génétique.
- **Divisive** : est une approche descendante où toutes les observations commencent dans un seul cluster, et des divisions sont effectuées de manière récursive en descendant dans la hiérarchie. L'approche divisive est utilisée pour la détection d'anomalies, la classification biologique, le traitement du langage naturel.

Critère de fusion : Il y a plusieurs critères de fusions qui sont possibles, comme par exemple le maximum (complete-linkage clustering) ou le minimum (single-linkage clustering).

Avantages CAH :

- **Robustesse aux formes complexes des données** : Le CAH peut être efficace pour identifier des structures complexes dans les données, y compris les clusters de formes non conventionnelles.
- **Interprétabilité** : Comme le CAH produit une structure hiérarchique, elle peut fournir des informations sur les relations de similarité entre les différents clusters à différents niveaux d'agrégation, ce qui peut être utile pour l'interprétation des résultats.
- **Pas de présupposition sur le nombre de clusters** : Le CAH ne nécessite pas de spécifier le nombre de clusters à l'avance, ce qui le rend plus flexible dans certaines situations où le nombre de clusters n'est pas connu à priori.

Inconvénients :

L'algorithme a une complexité assez élevée c'est à dire qu'elle vaut n^3 en temps et en espace n^2 .

3.1.3. Fuzzy

Le Fuzzy Clustering[W3] est une méthode très semblable à l'algo K-means. Il a été notamment utilisé dans une méthode améliorée pour le clustering d'ECG [KMKSY23]. On considère n individus, m classes, le score d'appartenance $wk(i)$ que l'individu i appartient à la classe k . Ce score est normalisé entre 0 et 1. L'algorithme le plus répandu est le fuzzy C-means :

- Initialiser un m nombre de classe.
- Assigner le score d'appartenance aléatoire à tous les individus.
- Répéter jusqu'à convergence de l'algo :
 - Calculer le centroïde de chaque cluster (moyenne pondérée par $wk(i)$).
 - Recalculer le score d'appartenance.

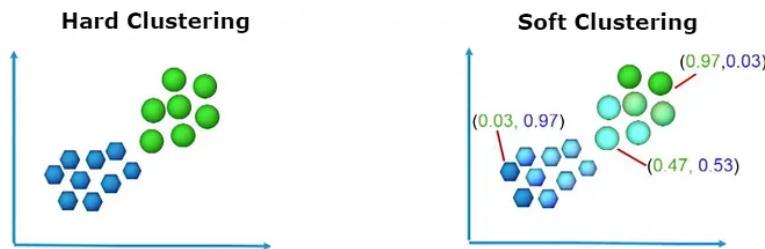


Figure 3 : Comparaison entre le clustering Fuzzy(soft) et le classique(hard) (from [medium](#))

Avantage Fuzzy C-means :

- Résistance aux outliers et aux bruits.
- Bonne classification pour les données superposées, c'est-à-dire proches mais associées à différents clusters.

Inconvénients :

- Complexité plus grande que K-means impliquant un temps d'exécution plus long.
- Nécessite l'initialisation comme K-means.

3.1.4. Spectral

Le Clustering Spectral [W4] est une technique qui transforme la donnée d'entrée via les vecteurs propres d'une matrice laplacienne [W5], puis réalise la classification sur une nouvelle matrice de donnée. Voici le procédé :

1. Générer un graphe G non-orienté, avec D sa matrice des degrés associé et A sa matrice d'adjacence
2. Calculer la matrice Laplacienne selon : $L = D - A$ (cf. figure 11)
3. Former la matrice X à partir des k premiers vecteurs propres en colonnes.
4. Réaliser un clustering au choix sur la matrice X

Cette méthodologie permet de Réduire la dimension de n variables à k variables où k est le nombre de clusters souhaité. Celà pourrait se révéler intéressant pour la vue génétique qui à une grande dimension [XYWSZWZ23].

Avantage :

- Méthode permettant de réduire la dimension avant un clustering.
- Réduire les erreurs des outliers et du bruit.

Inconvénients :

- Coût supplémentaire en empreinte mémoire et temps.

3.2. Multi-view learning

Le multi-view learning est une méthode d'apprentissage où les données sont représentées par plusieurs vues différentes. Dans notre cas les vues sont les données présentées dans la partie *description de la données*. Nous allons présenter quelques méthodes couramment utilisées.

3.2.1. Co-training

C'est une approche semi-supervisée ce qui signifie qu'il faut des données labellisées que nous n'avons pas cependant nous allons quand même décrire son fonctionnement brièvement. Le Co-training est l'une des premières méthodes d'apprentissage multi-vues. Son principe est le suivant : Il s'entraîne en alternance pour maximiser l'accord mutuel sur deux vues distinctes des données non étiquetées.

Le succès des algorithmes de co-training repose principalement sur trois hypothèses :

- la suffisance : chaque vue est suffisante pour qu'un algorithme de clustering soit possible sur elle-même.
- compatibilité : la fonction cible des deux vues prédit les mêmes labels pour les caractéristiques co-occurentes avec une forte probabilité.
- indépendance conditionnelle : les vues sont conditionnellement indépendantes.

L'hypothèse d'indépendance conditionnelle est critique, mais elle est généralement trop forte pour être satisfaite en pratique. Plusieurs alternatives plus faibles ont été considérées par la suite. Une image d'illustration est présentée en annexe.(cf. figure 9)

3.2.2. Multiple Kernel Learning (MKL)

Multiple Kernel Learning est créé pour contrôler la capacité de recherche des noyaux de matrice. Cependant il est aussi utilisé pour résoudre des problèmes de données multi-vues.

Les algorithmes d'apprentissage pour ces méthodes ont été développés aussi bien pour l'apprentissage non supervisé, semi-supervisé et supervisé. Plusieurs algorithmes de clustering du noyau ont été développés pour l'apprentissage non supervisé, semi-supervisé et supervisé.

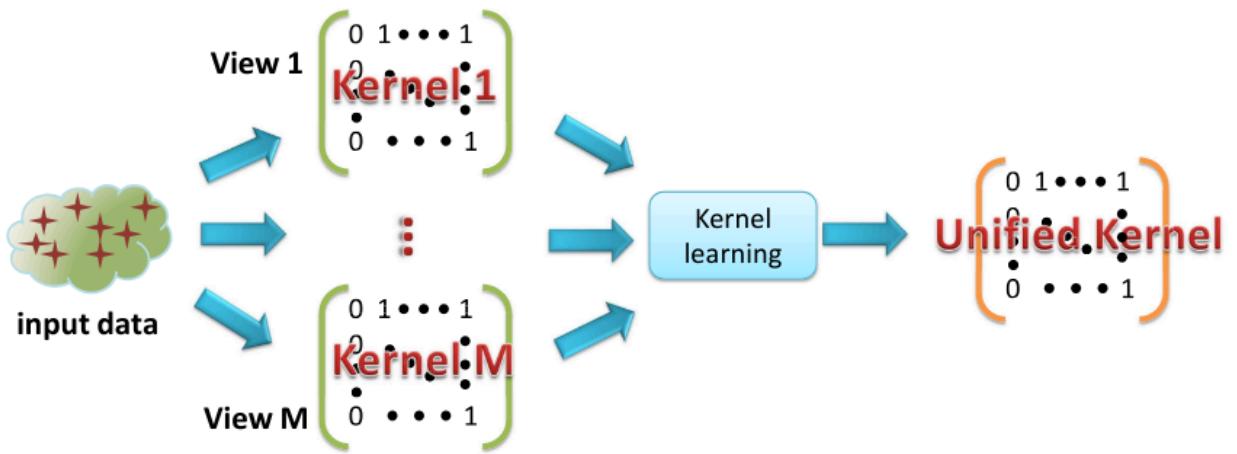


Figure 4 : Sketch map of multiple kernel learning. (From [XTX13])

3.2.3. Subspace learning

Le Subspace learning vise à découvrir les sous-espaces de faibles dimensions dans les données avec beaucoup de dimensions. Cette catégorie d'apprentissage est bien connue, en effet elle comprend en autre l'Analyse en composantes principales (PCA). Dans le contexte multi-vues, cette méthode est utilisée pour des sous-espaces communs ou partagés parmi les différentes vues.

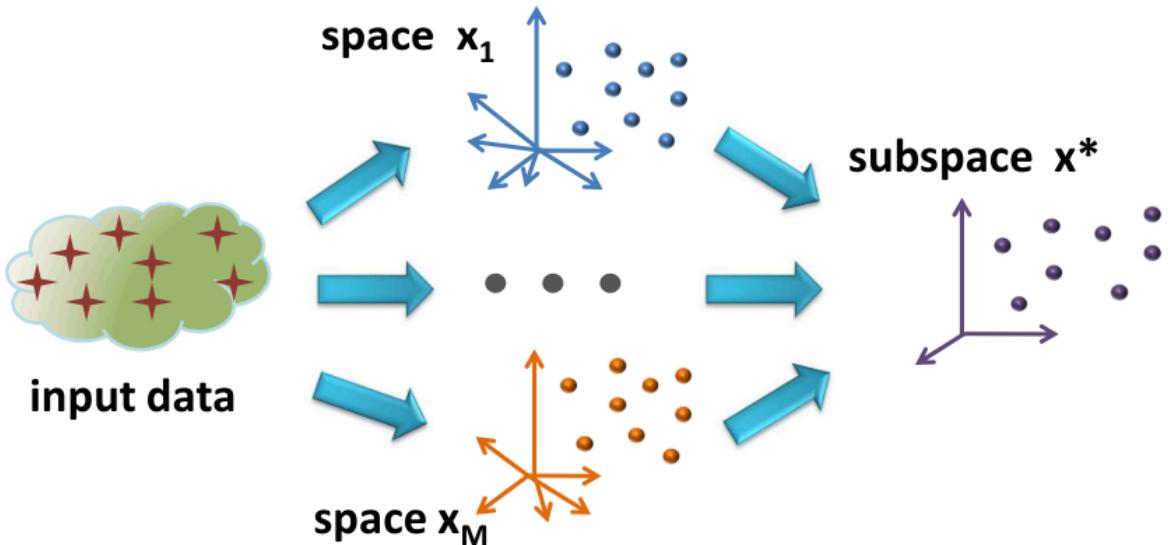


Figure 5 : Sketch map of subspace learning for multi-view data. (From [XTX13])

3.3. Multi-view clustering

En ce qui concerne les méthodes de clustering multi-vues, notre état de l'art n'est pas encore fini. Nous allons présenter les méthodes que nous avons assimilées, et énumérer quelques méthodes qui méritent plus de documentations afin qu'elles soient comprises.

Méthodes assimilées :

Nous avons la méthode MVC-LFA qui signifie Multi-view Clustering via Late Fusion Alignment Maximization.[WLZTLHXY19]

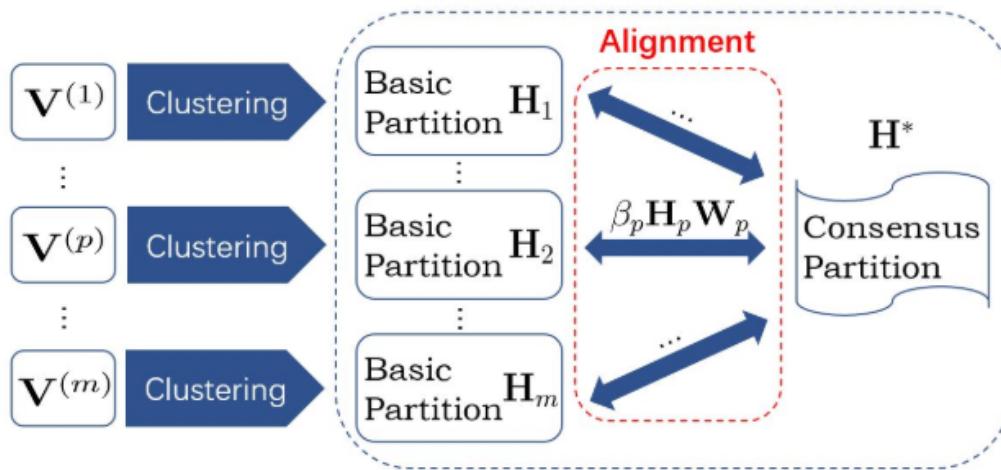


Figure 6 : Model pipeline for the proposed Multi-view Clustering via Late Fusion Alignment (MVC-LFA). (From [WLZTLHXY19])

La première étape est de faire un k-means qui donne les “Basic partition” de l'image. La seconde étape est l'alignement entre les partitions de bases pondérées générées.

L'algorithme de cette méthode réduit considérablement la complexité temporelle de $O(n^3)$ à $O(n)$. L'algorithme en question est présent en annexe. (cf. figure 10)

Nous avons aussi la méthode MKKM-MiR qui signifie Multiple Kernel K-means Clustering with Matrix-induced Regularization [LDYWZ16]. Cette méthode de clustering suit la méthode de Multiple Kernel Learning c'est-à-dire que chaque vue est représentée par une matrice de similarité. On applique un clustering k-means sur chaque vue séparément. Pour regrouper les partitions tout en conservant une cohérence entre celles-ci, on a besoin de mesurer la corrélation entre 2 kernels. Cette optimisation permet d'éviter que 2 kernels fortement corrélé soit sélectionné et ainsi réduit la redondance pour la construction du consensus. Cela favorise la diversité des kernels choisis grâce à cette régulation basée sur les matrices. L'algorithme converge en moins de 10 itérations et promet des performances qui surclasse les précédentes qui ne prenait pas assez en compte le lien entre les noyaux.

Méthode nécessitant approfondissement :

- PAC-Bayes
- A-MKKM Average multiple kernel k-means
- SB-KKM Single best kernel k-means
- LMKKM Localized multiple kernel k-means
- RMKKM Robust multiple kernel k-means [DZSWFWS15]
- CRSC Co-regularized spectral clustering
- Algorithmes de deep multi-view clustering

3.4. Evaluation et métrique des modèles

L'évaluation des modèles de clustering non supervisée est primordial. On définit un bon clustering comme satisfaisant si il respecte 3 axiomes [PB19] :

1. **Scale Invariance** : invariant si on multiplie la distance entre les données.
2. **Richness** : toute partition de la donnée est un résultat possible du clustering.
3. **Consistency** : invariant si la distance intra-classe diminue et/ou la distance inter-classe augmente.

Il a été démontré que un algo de clustering ne peut satisfaire que au plus 2/3 de ces règles.

Grâce à des métriques de performance on peut déterminer quelle approche est la plus optimale. Parmi ces métriques il y a celle qui dépendent d'une matrice de confusion[W6] : **Accuracy, Precision, Recall, F-score et métrique combinatoire**, utilisées en apprentissage supervisé. Dans notre cas, nous n'avons pas de labels dans notre données, alors nous ne pourrons pas profiter de cette évaluation externe.

Nous allons donc utiliser exclusivement des validation interne pour évaluer les modèles. Il y a 2 façons de mesurer l'efficacité de notre clustering : la cohésion intra-classe à maximiser et la séparation inter-classe à minimiser. Certaine métrique combine cohésion intra-classe et séparation inter-classe, c'est le cas de : **index Dunn, silhouette coefficient, Xie-Beni score**.

- Index Dunn fait le ratio entre la plus petite distance inter-classe et la plus grande distance intra-classe. Ce score est à maximiser
- Silhouette coefficient, compris entre -1 et 1 est une métrique très populaire. Une valeur positive indique une bonne séparation des données, une valeur négative indique des clusters trop superposés et enfin proche de 0 la valeur indique que la donnée est répartie uniformément. Sa complexité est grande ce qui la rend compliquée pour les très gros volumes de données.
- Xie-Beni est une métrique construite pour le fuzzy clustering mais est aussi applicable au autre clustering.

4. Conclusion

Dans ce rapport nous avons abordé le sujet du clustering multi-vues non supervisé. Nous avons fait un état de l'art en 3 parties résumant les méthodes et algorithmes de clustering, puis des nouvelles méthodes de multi-view learning/clustering.

L'étude de l'état de l'art nous a permis de découvrir le domaine du multi-vues. Ce sujet, riche et complexe, a suscité notre intérêt et a stimulé notre curiosité intellectuelle. Il nous a même amenés à nous interroger sur des problématiques plus spécifiques comme celle de l'*incomplete multi-view clustering*.

L'exploration de ces thématiques nous a donné l'opportunité de mettre un pied dans le monde de la recherche.

Nous sommes impatients de poursuivre notre travail sur ce projet, comprenant la fin de l'état de l'art ainsi qu'une comparaison approfondie de chaque méthode, en s'appuyant sur les évaluations et métriques qui sera réalisée en parallèle de nos propres reproductions applicatives sur le jeux de données fournis. Ces méthodes prometteuses permettront dans notre travail futur de résoudre la problématique d'extraction de profils de patients à partir de données de cardiomyopathie héréditaire.

Références

- [W1] <https://fr.wikipedia.org/wiki/Cardiomyopathie>
- [IEAAJ23] Abiodun M. Ikotun, Absalom E. Ezugwu, Laith Abualigah, Belal Abuhaija, Jia Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data", 2023
- [W2] https://en.wikipedia.org/wiki/Hierarchical_clustering
- [W3] https://en.wikipedia.org/wiki/Fuzzy_clustering
- [KMKSY23] Sanjay Kumar, Abhishel Mallik, Akshi Kumar, Javier Del Ser, Guan Yang, "Fuzz-ClustNet: Coupled fuzzy clustering and deep neural networks for Arrhythmia detection from ECG signals", 2023
- [W4] https://en.wikipedia.org/wiki/Spectral_clustering
- [W5] https://en.wikipedia.org/wiki/Laplacian_matrix
- [W6] https://en.wikipedia.org/wiki/Confusion_matrix
- [XYWSZWZ23] Zhaozhao Xu, Fangyuan Yang, Hong Wang, Junding Sun, Hengde Zhu, Shuihua Wang, Yudong Zhang, "CGUFS: A clustering-guided unsupervised feature selection algorithm for gene expression data", 2023
- [PB19] Julio-Omar Palacio-Nino, Fernando Berzal, "Evaluation Metrics for Unsupervised Learning Algorithms", 2019
- [XTX13] Chang Xu, Dacheng Tao, Chao Xu, "A Survey on Multi-view Learning", 2013
- [GM14] Mehmet Gonen, Adam A. Margolin, "Localized Data Fusion for Kernel k-Means Clustering with Application to Cancer Biology", 2014
- [WLZTLHXY19] Siwei Wang, Xinwang Liu, En Zhu, Chang Tang, Jiyuan Liu, Jingtao Hu, Jingyuan Xia, Jianping Yin, "Multi-view Clustering via Late Fusion Alignment Maximization", 2019
- [LDYWZ16] Xinwang Liu, Yong Dou, Jianping Yin, Lei Wang, En Zhu, "Multiple Kernel k-Means Clustering with Matrix-Induced Regularization", 2016
- [ZXXS17] Jing Zhao, Xijiong Xie, Xin Xu, Shiliang Sun, "Multi-view learning overview: Recent progress and new challenges", 2017
- [DZSWFWS15] Liang Du, Peng Zhou, Lei Shi, Hanmo Wang, Mingyu Fan, Wenjian Wang, Yi-Dong Shen, "Robust multiple kernel k-means", 2015

Annexe

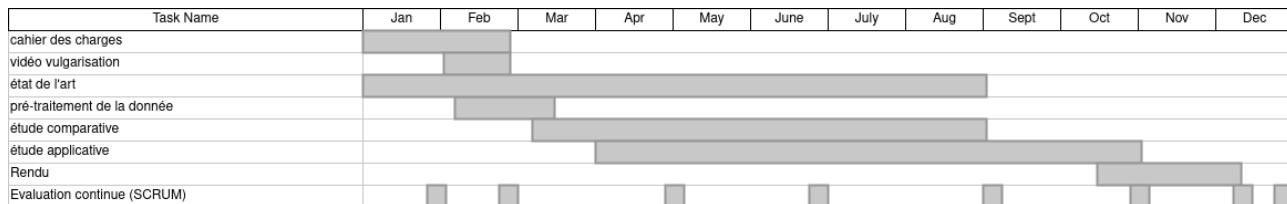


Figure 7 : Diagramme de Gantt

Domaine d'amélioration	Méthode
Automatic Specification of K	Learning K using a randomly sampled dataset Unsupervised learning schema Learning schema using entropy-regularization terms The use of PCA Optimization of BIC or AIC Statistical test on cluster centers One way analysis of variance Atkinson index
Improved Initial Centroids Selection	Atkinson index Datapoint's probability contribution Sampling more than k centers Cluster elimination and division Gaussian distance ratio for dissimilarity evaluation Heuristic approach Pairs with the longest distance Dynamic adding of cluster center Selection rule Consistent similar points between pairs Standardization of data for the center of gravity Using negative and positive attributes values Locality sensitivity on subtractive clustering Using the median of data objects Optimization of initial cluster centers The use of medoids
The use of PCA for Cluster Centroids	New lower bound for K-means objective function Adaptive search of GA Using PCA and division into percentiles for efficient initial coordinates for centroids
Resolution of local optimal convergence	Spectral relaxation Maximum triangle rule Selection and erasure rule Additive perturbation stability instances

Figure 8 : Exemple non exhaustif d'Amélioration de Kmeans

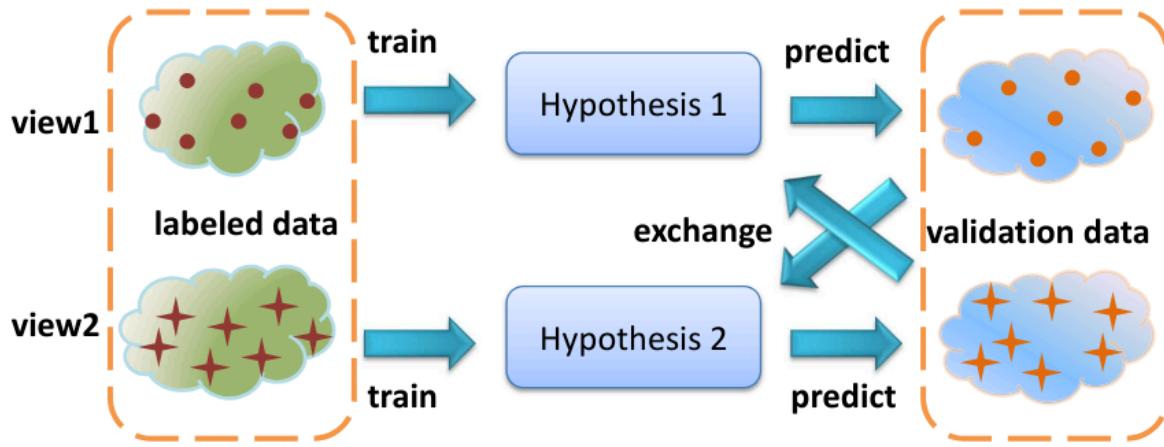


Figure 9 : The process of co-training style algorithms. (From [XTX13])

Algorithm 1 Multi-view Clustering via Late Fusion Alignment Maximization

Input: $\{\mathbf{H}_p\}_{p=1}^m$, k , λ and ϵ_0 .

Output: \mathbf{H}^* and β .

- 1: Initialize $\{\mathbf{W}_p\}_{p=1}^m = \mathbf{I}_k$, $\beta = \frac{1}{\sqrt{m}}$ and $t = 1$.
 - 2: Calculate \mathbf{M} by kernel k -means with average kernel.
 - 3: **while** not converged **do**
 - 4: Update \mathbf{H}^* by solving Eq. (7) with fixed $\{\mathbf{W}_p\}_{p=1}^m$ and β .
 - 5: Update $\{\mathbf{W}_p\}_{p=1}^m$ with fixed \mathbf{H}^* and β by solving Eq. (10).
 - 6: Update β by solving Eq. (12) with fixed \mathbf{H}^* and $\{\mathbf{W}_p\}_{p=1}^m$.
 - 7: $t = t + 1$.
 - 8: **end while** $(\text{obj}^{(t-1)} - \text{obj}^{(t)}) / \text{obj}^{(t)} \leq \epsilon_0$
 - 9: **return** \mathbf{H}^* and β .
-

Figure 10 : MVC-LFA algorithm with equation. (From [WLZTLHXY19])

Labelled graph	Degree matrix	Adjacency matrix	Laplacian matrix
	$\begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ -1 & -1 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}$

Figure 11 : Exemple de Calcul de la matrice Laplacienne from wikipedia