L'apprentissage non-supervisé et ses contradictions

Jérémie Sublime

L'apprentissage non-supervisé et ses contradictions

Jérémie Sublime [1, 2]

L'apprentissage machine, également appelé apprentissage artificiel [2], ou encore *machine learning* en anglais, est un domaine scientifique lié à l'intelligence artificielle et qui se trouve à la frontière entre l'informatique et les statistiques. L'objectif de ce domaine est la conception et l'étude d'algorithmes capables « d'apprendre » à partir de données. Ces algorithmes d'apprentissage sont ensuite utilisés pour des tâches diverses et variées : analyse et traitement d'images (i.e. reconnaissance de visages), détection d'anomalies (i.e. connexions frauduleuses), systèmes de recommandation (Google, Youtube, Amazon, Netflix, etc.), tri et catégorisation d'information, et même pour apprendre à des programmes à jouer aux échecs, au Go [3], ou aux jeux vidéos [4].

On divise généralement l'apprentissage machine en trois grandes catégories : l'apprentissage supervisé, non-supervisé et par renforcement.

Dans l'apprentissage supervisé, les algorithmes sont entraînés à reconnaître et catégoriser des données à partir d'exemples étiquetés. On apprendra, par exemple, à un réseau de neurones à faire la différence entre des photos de chats ou de chiens en lui montrant des centaines d'images et en lui précisant à chaque fois de quel animal il s'agit de sorte qu'il finira par apprendre ce qui caractérise un chat ou un chien, et saura reconnaître ces caractéristiques sur une photo qu'il n'aura jamais vue.

1. ISEP, École d'ingénieurs du numérique.
2. LIPN - CNRS UMR 7030, LaMSN - Université Sorbonne Paris Nord.
3. https://www.bbc.com/news/technology-35785875.
4. DeepMind/AlphaStar.

DOI:10.48556/SIF.1024.19.145 **1024** – Bulletin de la SiF, numéro 19, Avril 2022, pp. 145–156

Dans l'apprentissage non-supervisé, conçu à la base comme une tâche exploratoire, les algorithmes sont amenés à traiter des données non-étiquetées pour y trouver des structures, des groupes d'objets similaires, ou des anomalies. Si on reprend l'exemple précédent avec les photos de chats et de chiens, on fournira cette fois-ci toutes les photos à l'algorithme sans lui préciser de quel animal il s'agit. Avec un peu de chance, la méthode non-supervisée s'apercevra toute seule qu'il y a deux animaux différents selon les photos, et fera un tri pour séparer les chiens et les chats. Mais il se peut aussi qu'elle décide de les trier par couleur (en mélangeant chiens et chats), qu'elle fasse de multiples catégories par race, ou encore qu'elle fasse un tri selon des critères difficilement compréhensibles pour un humain (la couleur du papier peint par exemple) en ne se préoccupant nullement des animaux présents sur les images. On parle de *clustering* lorsque l'apprentissage non-supervisé regroupe ensemble des données considérées comme similaires selon des critères plus ou moins transparents. Les groupes ainsi formés sont appelés *clusters*.

L'apprentissage par renforcement est un domaine un peu différent des précédents puisqu'on cherche ici à influencer le comportement d'un algorithme d'intelligence artificielle à partir de récompenses si l'algorithme se comporte comme on le souhaite, et de pénalités dans le cas contraire. Ce type d'apprentissage est davantage utilisé pour apprendre aux algorithmes à jouer à des jeux (en choisissant les bons coups à jouer) plutôt que pour traiter des données.

L'apprentissage supervisé est sans doute la catégorie la plus connue des chercheurs tous domaines confondus, avec ses réseaux de neurones aux performances extraordinaires en analyse d'image. L'apprentissage par renforcement, quant à lui, est le plus médiatisé auprès du grand public pour ses succès lorsque les machines battent les meilleurs humains sur des jeux ou des tâches que l'on estimait il y a quelques années encore trop complexes pour qu'un algorithme puisse vaincre un jour un humain. Quant à moi, je travaille depuis le début de ma carrière de chercheur sur l'apprentissage non-supervisé, la branche moins connue de l'apprentissage, la plus mal posée aussi, et qui sera le sujet de cet article. J'aborderai un certain nombre de contradictions et de questions que l'on peut voir émerger dans ce domaine un peu particulier.

Un apprentissage non-supervisé détourné de son but exploratoire initial ?

Lorsqu'il est présenté à des étudiants dans l'enseignement supérieur, l'apprentissage non-supervisé est généralement abordé sous l'angle du *clustering* : cette tâche exploratoire que j'ai évoquée plus haut et qui consiste à regrouper ensemble des données similaires dans des *clusters*. Ce cours sur le *clustering* fait généralement partie d'un module plus large d'apprentissage machine et est l'occasion de présenter les principales familles d'algorithmes de *clustering* : on présentera les K-moyennes [11]

pour les méthodes dites « basées centroïdes », le *clustering* basé sur la densité des données dans l'espace [5, 1], le *clustering* hiérarchique [16] qui construit les clusters sur plusieurs niveaux comme un dendrogramme, et parfois la version statisticienne des K-moyennes sous la forme des modèles de mixtures gaussien. Pour chaque famille d'algorithmes, on détaillera les grands principes, la complexité algorithmique, mais aussi quels types de clusters peuvent et ne peuvent pas être trouvés par ces algorithmes. Les structures que chacun des algorithmes peuvent identifier sont généralement illustrées comme dans la figure 1 ci-après (qui évoque aussi les temps de calcul). En effet, chaque famille d'algorithmes de *clustering* présuppose des formes de clusters à trouver et leur séparabilité.

C'est ce dernier point qui vient généralement mettre le premier coup de canif au concept de l'apprentissage non-supervisé comme méthode exploratoire : l'idée de devoir donner à l'avance presque toutes les informations sur les clusters qu'on va chercher si on veut avoir une petite chance de les trouver.

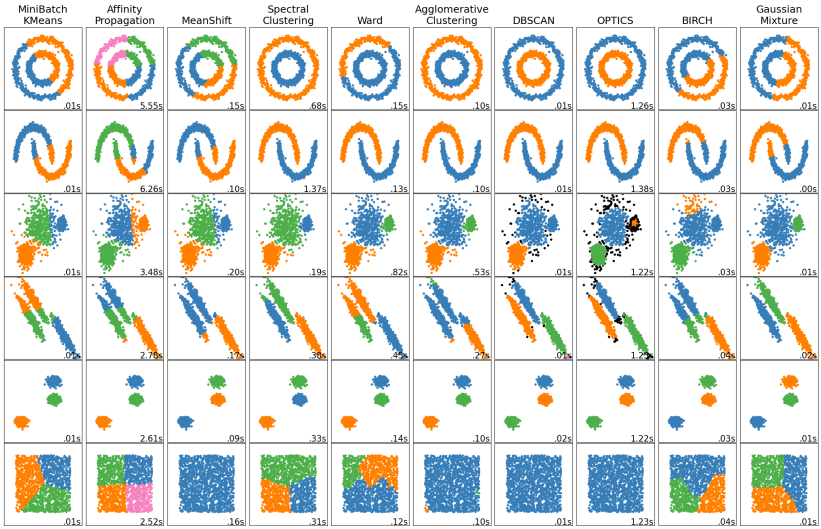

FIGURE 1. Quelques exemples de méthodes de clustering appliquées à diverses configurations de points dans l'espace. © Scikit-learn [5]

Certains réfutent parfois ce point en expliquant qu'il s'agit — avec le choix de l'algorithme — de faire une hypothèse a priori sur cette forme, et non pas de la

5. https://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_comparison.html#sphx-glr-auto-examples-cluster-plot-cluster-comparison-py.

connaître à l'avance. Cet argument est à mon avis naïf, étant donné l'impossibilité d'émettre une hypothèse éclairée sans avoir déjà une très bonne idée de ce qu'on cherche. Il souligne cependant bien le côté « problème mal posé » du clustering.

Si on regarde maintenant les applications concrètes de l'apprentissage non-supervisé dans la vie d'un chercheur ou d'un industriel, on s'aperçoit assez vite que dans la plupart des cas pratiques ce n'est pas tellement le côté exploratoire qui motive son utilisation, mais plutôt l'absence de données annotées. En effet, l'immense majorité des tâches de traitement de données ou d'images qui utilisent des techniques d'intelligence artificielle dans la vie réelle, sont des tâches pour lesquelles on cherche des classes précises et connues : dans mon exemple d'introduction, on voulait séparer les chiens et les chats, en imagerie médicale, on voudra par exemple classifier si des tissus sont cancéreux ou non, en imagerie satellite on voudra identifier et classer différents types de bâtiments vus du ciel, en marketing digital on voudra identifier des catégories sociaux-économiques bien précises pour cibler des publicités en ligne, etc. Cependant, il se trouve que ces domaines sont parfois confrontés à une absence de données annotées (absence totale ou nombre insuffisant) pré-existantes pour une application précise, empêchant alors d'avoir recours aux classiques algorithmes d'apprentissage supervisés. Dans ces conditions, l'utilisation de l'apprentissage non-supervisé — qui lui ne nécessite pas de données annotées — peut sembler être une alternative séduisante par rapport au coût de création d'une base suffisante de données étiquetées. On se retrouve alors de fait avec une utilisation détournée de l'apprentissage non-supervisé, qui a perdu tout caractère exploratoire, et est utilisé par défaut pour pallier un manque de données étiquetées. Et ce détournement n'est pas sans conséquences puisque ces algorithmes n'ont absolument pas été conçus pour que leurs clusters convergent comme par magie vers les classes recherchées pour les applications réelles. Les résultats sont souvent assez décevants, surtout lorsqu'ils sont comparés aux performances qu'on aurait pu avoir avec des étiquettes et un bon algorithme supervisé.

Mais alors pourquoi continue-t-on de détourner les méthodes non-supervisées plutôt que d'étiqueter des données pour passer ensuite au supervisé ? On peut citer deux explications complémentaires : la première est la relative simplicité des données et des méthodes, jusqu'à la fin des années 90. Cela permettait aux méthodes non-supervisées de conserver des performances correctes dans le meilleur des cas, et de pouvoir envisager un étiquetage manuel pas trop coûteux en cas de nécessité de passer à du supervisé. Dans le même temps, la complexité des données a également explosé, creusant ainsi les écarts de performances entre les algorithmes. La seconde explication est la percée récente des réseaux de neurones profonds qui sont particulièrement gloutons en données étiquetées pour pouvoir atteindre les excellentes performances que nous leur connaissons. Si je résume, nous sommes donc dans une conjoncture où les données sont à la fois plus nombreuses, plus complexes et donc

plus difficile à étiqueter, et nous avons aussi des algorithmes supervisés qui nécessitent de plus en plus de données annotées.

L'imagerie satellite et l'imagerie médicale — deux domaines que je connais bien — sont de bons exemples de l'apparition de plus en plus fréquente de ce type de problème. En imagerie satellite, la variété des paysages, des capteurs (dont la résolution et le nombre de bandes varient) et des applications, fait que des données annotées pour une application seront rarement réutilisables pour un autre problème avec des images d'une autre région. En imagerie médicale, on retrouve ce même problème de format d'acquisition qui change et de variété des applications (rendant les annotations déjà faites pas forcément utiles d'un problème à l'autre), mais aussi des conflits d'experts qui ne sont pas d'accord sur la façon dont certaines images complexes devraient être annotées, ajoutant donc en plus un problème de confiance dans les annotations. Ainsi, lorsque ces experts métier (géographes, professeurs de médecine, ou autres) se tournent vers les spécialistes de l'apprentissage supervisé pour voir si on ne pourrait pas automatiser l'interprétation de leurs images, ou même prédire de futures évolutions avec le dernier réseau de neurones à la mode, on leur répond souvent que c'est possible mais qu'il faudra fournir des centaines (parfois des milliers) d'images précisément annotées. Si les très gros instituts et conglomérats peuvent peut-être fournir ces annotations (et parfois les partager par la suite avec le reste de la communauté), les autres doivent renoncer à l'idée ou tenter leur chance avec un spécialiste de l'apprentissage non-supervisé.

L'apprentissage non-supervisé a fait beaucoup de progrès depuis ses débuts, et est lui aussi entré dans l'ère des réseaux de neurones. Il existe donc quelques réponses dans l'arsenal non-supervisé qui permettent de s'attaquer à des images ou des données complexes avec des outils adaptés : l'auto-encodeur [9, 7] reste à mon sens l'architecture de base la plus utilisée lorsqu'on veut combiner apprentissage non-supervisé et réseaux de neurones. Cette architecture qui est présentée dans sa forme la plus simple sur la Figure 2 ci-après, vise à compresser et à reconstruire les données en entrées via un encodeur et un décodeur symétriques qui pourront être composés de diverses couches de neurones qui dépendront de l'application. L'entonnoir de l'auto-encodeur permet généralement de récupérer une version des données débruitée et encodée dans un espace latent qui sera plus facilement utilisable, par exemple par une méthode de clustering classique. Ainsi, l'auto-encodeur permet à la fois de réutiliser la majorité des composants (couches et fonctions d'activation) des réseaux de neurones classiques, et peut-être combiné soit à des méthodes de clustering classiques, soit à quelques couches neuronales supplémentaires spécifiques pour obtenir l'algorithme non-supervisé voulu.

Il n'en reste pas moins que malgré la possibilité technique de construire des réseaux de neurones non-supervisés capables de s'attaquer à des données aussi complexes que des images satellites ou médicales, nous restons dans ce cas de figure, pas du tout exploratoire, où l'on attend d'une méthode non-supervisée qu'elle trouve des

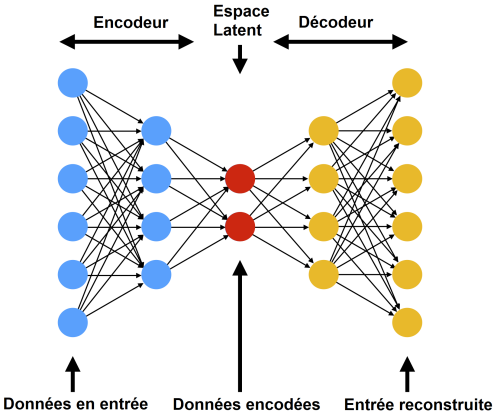

FIGURE 2. Exemple d'une architecture de base d'auto-encodeur qu'on pourra modifier selon l'application.

clusters qui vont coller avec des classes réelles très précises sur lesquelles les experts eux-même ne sont parfois pas d'accord. Pour prendre un exemple issu de mon manuscrit d'habilitation [15] : "*Attendre d'un réseau de neurones non-supervisé qu'il vous trouve des lésions sur une image médicale, c'est comme demander à un enfant de 5 ans de colorier ce qui lui* ***paraît remarquable*** *sur cette même image. Les deux ont à peu près les mêmes capacités visuelles, le réseau de neurones sera peut-être légèrement meilleur en coloriage, et aucun des deux n'a la moindre idée de ce qui est sur l'image ou de ce que vous attendez de lui. Vous aurez votre coloriage, mais il n'est pas certain que le résultat soit le résultat attendu.*"

Dans ces conditions, et puisque nous assumons que nous ne sommes plus dans un contexte exploratoire mais bien dans un cas où l'on sait ce qu'on cherche, où l'on a juste pas assez de données annotées : puisque l'on possède quelques données annotées, pourquoi ne pas les utiliser pour aider l'algorithme de clustering ? Ce cadre a d'ailleurs été étudié et s'appelle l'apprentissage semi-supervisé. Hélas, les choses ne sont pas si simples. Si effectivement l'apprentissage semi-supervisé est bien un domaine qui existe et qui consiste à fournir des données étiquetées et d'autres non-étiquetées à un algorithme d'apprentissage pour qu'il apprenne mieux, pour le moment, il concerne principalement l'amélioration des performances d'algorithmes supervisés auxquels il faudra quand même fournir un maximum de données étiquetées, et pour lesquels fournir quelques données non-étiquetées pourra améliorer les performances en fin d'apprentissage. En revanche, les expériences pour faire l'inverse — avoir majoritairement des données non-étiquetées et fournir quelques données étiquetées à un algorithme non-supervisé (comment ?) — sont encore rares et très balbutiantes. On aura donc tendance à considérer que le semi-supervisé n'existe

pas vraiment, ou reste en tout cas très supervisé.

De cette section, on pourrait conclure que non seulement les méthodes non-supervisées sont finalement assez peu utilisées dans un cadre vraiment exploratoire, mais surtout que même quand on n'a pas d'autre choix que de les utiliser il y a de fortes chances que les résultats soient décevants. Cependant, nous allons voir par la suite qu'on peut tout de même utiliser les algorithmes non-supervisés avec une certaine efficacité, à condition de bien appréhender les limites de la non-supervision.

L'apprentissage non-supervisé serait-il quand même un peu supervisé ?

Étant donné ce qui a été évoqué précédemment sur les difficultés des méthodes non-supervisées à répondre aux attentes de trouver des classes réelles sur des applications pratiques, je me suis souvent posé la question de savoir comment d'autres chercheurs et moi-même arrivions tout de même à atteindre des résultats honorables (pour du non-supervisé) sur des problèmes complexes avec nos algorithmes. En effet, si on regarde la littérature scientifique, cet apprentissage non-supervisé — qui semble ne rien avoir pour lui — reste utilisé par de nombreux chercheurs dans de nombreux domaines, en particulier à cause du manque de données étiquetées. Et les résultats sont parfois impressionnants.

C'est un récent co-encadrement de stage et une rétrospection sur mon expérience personnelle qui m'ont apporté quelques éléments de réponse que je vais partager avec vous. Le stage en question avait pour but de faire de la segmentation de lésions liées à la DMLA[6] sur des images médicales en infrarouge [14] : un problème assez classique d'imagerie médicale donc, avec des lésions plus ou moins complexes à détourer selon les images, des experts pas d'accord sur les cas complexes, et pas assez d'images bien annotées pour entraîner un réseau de neurones supervisé de type UNet [12]. Notre stagiaire c'est donc tourné vers les WNets [17], le réseau de neurones équivalent aux UNets en non-supervisé pour faire de la segmentation. Ce réseau repose d'ailleurs en partie sur l'architecture d'auto-encodeur évoquée précédemment. Les premiers résultats furent ce qu'on pouvait attendre : mauvais. Les lésions qui nous intéressaient n'étaient que rarement séparées du reste, coupées en de multiples classes, ou complètement ignorées au profit d'autres éléments structurels présents dans les images et qui avaient sans doute été jugés plus remarquables par l'algorithme non-supervisé. Malgré tout, nous avons persévéré : notre stagiaire avait pour habitude de nous montrer de nouveaux résultats tous les 3 ou 4 jours au fil des modifications apportées sur les fonctions d'activation des neurones, les changements de fonctions de perte, ou de tous autres paramètres de son algorithme. Il s'est avéré qu'au fil des modifications et de nombreux essais et erreurs, les résultats se sont

6. Dégénérence maculaire liée à l'âge, une pathologie de l'œil.

lentement mais sûrement améliorés. Il m'est alors apparu comme assez évident que faute de pouvoir superviser un algorithme directement et automatiquement avec de grands volumes de données étiquetées, il était possible de le guider au fil des essais à partir de résultats visuels et des résultats sur les quelques données étiquetées à disposition : il est possible de choisir certains paramètres en fonction de ce qui fonctionne ou de ce qui ne fonctionne pas. Mais n'y a-t-il pas là une forme de supervision qui ne dit pas son nom ?

En effet, en y réfléchissant bien, le processus que nous avons décrit est plus ou moins le même que celui que j'ai pu constater chez la majorité des chercheurs qui font de l'apprentissage non-supervisé : on part d'une méthode de base (et nous avons vu que le choix de cette méthode va favoriser certaines formes de clusters et peut donc déjà être vu comme une forme de supervision), puis on va modifier et guider cette méthode par tous les moyens détournés possibles et imaginables jusqu'à avoir des résultats satisfaisants sur l'application qui nous intéresse. Et ce processus de guidage — quel que soit sa forme [6] — nécessite une certaine expertise et demande du temps ; en particulier par rapport aux méthodes supervisées pour lesquelles il suffit généralement d'adapter un peu les formats en entrée de l'algorithme pour que de bons résultats sortent assez rapidement. Je pense donc pouvoir affirmer que le secret d'un apprentissage non-supervisé qui donne de bons résultats est que tout le monde « triche » en compensant l'absence de supervision via des données étiquetées par une autre forme de supervision : celle d'un processus de guidage manuel de l'algorithme (du choix de la méthode de base aux paramètres) jusqu'à atteindre les résultats voulus.

Cette reconnaissance d'une supervision autrement que directement par les données en non-supervisé (a priori courante, mais rarement admise ouvertement) soulève tout de même quelques questions :

— peut-on toujours parler d'apprentissage non-supervisé dans ce contexte ?

— pourrait-on automatiser ce processus ? En effet, si on regarde de plus près, cela ressemble à de l'apprentissage par renforcement qui serait fait à la main. Mais, c'est également assez proche de l'autoML [8] tel qu'on le pratique en apprentissage supervisé ;

— quelles conséquences le recours à ce processus peut-il avoir lors de la comparaison des performances de plusieurs méthodes non-supervisées ?

Je vais m'attarder sur la dernière question qui est assez essentielle pour de nombreux chercheurs en apprentissage : nous sommes tous plus ou moins soumis à une certaine injonction à publier, et chacun sait que — à tort ou à raison — pour être publiable un algorithme applicatif doit bien souvent montrer qu'il fait mieux (ou au moins aussi bien) que ses concurrents de la littérature. Dès lors qu'on se place dans le prisme d'une application réelle, on peut oublier les indices classiques de clustering [13, 3], et on se concentrera sur une évaluation par le prisme des indices supervisés

classiques tels que la précision, le rappel ou le score F1. Bien que nous soyons en contexte non-supervisé ici, nous avons tout de même ces quelques images annotées à notre disposition qui, si elles ne sont pas assez nombreuses pour entraîner une méthode supervisée, suffisent généralement pour avoir une évaluation quantitative. Or, si on admet l'existence de ce processus de supervision détournée des méthodes non-supervisées pour les rendre plus compétitives, on s'aperçoit assez vite que la comparaison ne sera pas équitable entre une méthode nouvellement proposée par des chercheurs qui l'auront sur-optimisée et qui ne pourra que battre ses concurrents non-supervisés réutilisés depuis d'autres applications plus ou moins proches et qui eux n'auront pas forcément reçu le même niveau d'optimisation par une supervision détournée. Bien que n'ayant pas de solution à ce problème, je pense que ce sujet méritait d'être détaillé, d'autant que ce phénomène du nouvel algorithme qui écrase complètement des concurrents pourtant a priori sérieux (mais sans doute pris sur l'étagère) est assez courant dans la littérature non-supervisée.

A l'issue de cette seconde partie de réflexion, on s'aperçoit finalement que la clé d'un apprentissage non-supervisé qui donne des résultats corrects sur des données complexes est très certainement d'apporter la supervision autrement que par les étiquettes, de manière à tout de même guider l'algorithme. Si ce processus est aujourd'hui à mon avis massivement pratiqué, il reste très artisanal et serait à perfectionner. Est-ce à dire pour autant qu'en faisant cela, les méthodes non-supervisées peuvent faire aussi bien que les méthodes supervisées ? La réponse courte sera non. Certes cette supervision détournée améliore considérablement les résultats, mais elle

ne battra jamais une supervision directe et automatisée par les étiquettes des classes recherchées. Cependant, le gain de temps reste considérable par rapport au processus de construction à la main d'un corpus de données annotées de quelques milliers d'images.

Conclusion

Nous avons abordé dans ce texte plusieurs aspects et contradictions de l'apprentissage non-supervisé. Nous avons vu que contrairement à l'idée que l'on s'en fait et à la façon dont on l'enseigne, il n'est *a priori* que rarement exploratoire : on l'applique comme solution par défaut sur des problèmes bien supervisés pour lesquels on n'a pas ou peu de données étiquetées. En y regardant de plus près, on s'aperçoit aussi que ce qu'on appelle apprentissage non-supervisé est finalement assez souvent supervisé de manière détournée : que ce soit par le choix de l'algorithme qui va favoriser certains clusters constituant ainsi une première forme de supervision, ou bien dans des choix plus pointus d'architecture et de modèles qui seront des points encore plus forts de supervision.

Personnellement, je n'aime pas le qualificatif de « non-supervisé » qui est à mon avis trompeur. Je ne suis pas le seul. Certains préfèrent maintenant parler de *self-supervised learning* [18, 4, 10] : l'apprentissage des données par les données. C'est un meilleur nom. Et cette idée fonctionne très bien sur certaines tâches. Cependant, et bien que ce ne soit pas le sujet de ce texte, on pourrait argumenter que les principaux succès du *self-supervised learning* ont été accomplis en indiquant aux algorithmes quelle partie des données pourrait être apprise à partir de quelle autre, et donc avec une supervision externe. Je reste par conséquent peu convaincu que les données seules puissent jamais apporter la supervision nécessaire pour les nombreuses tâches de classification ou de traitement d'image avancées qu'on souhaite réaliser sans avoir à annoter des milliers de données manuellement. En effet, ces annotations manuelles servent à indiquer à l'algorithme quel est son but et quelles classes il doit trouver : elles servent à lui expliquer pourquoi il est intéressant de grouper par exemple des chiens et des chats plutôt que de grouper des photos majoritairement vertes. Aussi, il me semble que penser l'apprentissage non-supervisé sans concevoir qu'il y a un but derrière est illusoire. Une forme de supervision — peu importe laquelle, et qui pourra être plus ou moins complexe selon le but recherché — doit exister pour guider l'algorithme.

Au delà de la reconnaissance des forces et des limites de l'apprentissage non-supervisé, et d'un éventuel changement de nom pour mieux coller à la réalité de son utilisation, il me semble que les prochaines avancées pour sortir l'apprentissage non-supervisé de ses contradictions devraient se concentrer sur deux points évoqués dans cet article :

(1) le développement de méthodes *faiblement supervisées* où le guidage d'un algorithme vers des classes réelles pourrait se faire à partir de très peu d'exemples annotés en plus de données non-annotées ;

(2) la reconnaissance et la formalisation de l'actuel processus de supervision détournée, de manière à pouvoir l'utiliser plus efficacement et à assurer des comparaisons équitables de méthodes non-supervisées.

Références

[1] M. Ankerst, M. M. Breunig, H. Kriegel, and J. Sander. OPTICS : Ordering Points To Identify the Clustering Structure. In *In ACM SIGMOD international conference on Management of data*, pages 49–60. ACM Press, 1999.

[2] A. Cornuéjols and L. Miclet. *Apprentissage artificiel : Concepts et algorithmes*. Eyrolles, June 2010.

[3] D. L. Davies and D. W. Bouldin. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2) :224–227, Feb. 1979.

[4] C. Doersch and A. Zisserman. Multi-task self-supervised visual learning. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2070–2079, 2017.

[5] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, page 226–231. AAAI Press, 1996.

[6] P. Gançarski, T.-B.-H. Dao, B. Crémilleux, G. Forestier, and T. Lampert. Constrained Clustering : Current and New Trends. In P. Marquis, O. Papini, and H. Prade, editors, *A Guided Tour of AI Research*, volume 2. Springer, 2020.

[7] X. Guo, X. Liu, E. Zhu, and J. Yin. Deep Clustering with Convolutional Autoencoders. In D. Liu, S. Xie, Y. Li, D. Zhao, and E.-S. M. El-Alfy, editors, *Neural Information Processing*, pages 373–382. Springer International Publishing, 2017.

[8] I. Guyon, I. Chaabane, H. J. Escalante, S. Escalera, D. Jajetic, J. R. Lloyd, N. Macià, B. Ray, L. Romaszko, M. Sebag, A. R. Statnikov, S. Treguer, and E. Viegas. A brief review of the chalearn automl challenge : Any-time any-dataset learning without human intervention. In F. Hutter, L. Kotthoff, and J. Vanschoren, editors, *Proceedings of the 2016 Workshop on Automatic Machine Learning, AutoML 2016, co-located with 33rd International Conference on Machine Learning (ICML 2016), New York City, NY, USA, June 24, 2016*, volume 64 of *JMLR Workshop and Conference Proceedings*, pages 21–30. JMLR.org, 2016.

[9] G. E. Hinton and R. R. Salakhutdinov. Reducing the Dimensionality of Data with Neural Networks. *Science*, 313(5786) :504–507, 2006.

[10] Y. LeCun. Self-supervised learning. In *Plenary talk at the 8th International Conference on Learning Representations, ICLR 2020, Online*, 2020.

[11] J. B. MacQueen. Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.

[12] O. Ronneberger, P. Fischer, and T. Brox. U-Net : Convolutional Networks for Biomedical Image Segmentation. volume 9351, pages 234–241, 10 2015.

[13] R. Rousseeuw. Silhouettes : a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics.*, 20 :53–65, 1987.

[14] C. Royer, J. Sublime, F. Rossant, and M. Pâques. Unsupervised approaches for the segmentation of dry ARMD lesions in eye fundus cslo images. *J. Imaging*, 7(8) :143, 2021.

[15] J. Sublime. *Contributions to modern unsupervised learning : Case studies of multi-view clustering and unsupervised Deep Learning. (Contributions à l'apprentissage non-supervisé moderne : Applications aux cas du clustering multi-vue et de l'apprentissage profond non-supervisé)*. 2021.

[16] J. H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301) :236–244, 1963.

[17] X. Xia and B. Kulis. W-net : A deep model for fully unsupervised image segmentation. *CoRR*, abs/1711.08506, 2017.

[18] X. Zhai, A. Oliver, A. Kolesnikov, and L. Beyer. S4l : Self-supervised semi-supervised learning. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1476–1485, 2019.