Méta apprentissage pour un Apprentissage Intelligent

Laboratoire : LIPN-UMR 7030 Université Sorbonne Paris Nord

Direction de thèse : Aomar OSMANI

Nom de la candidate : Meliha AKOUBA

Mots Clés: Machine Learning Theory, Meta-Learning, Structural Risk Minimisation, Hilbert space, Lie Groups, weak and strong convergences, probabilistic theory, Explanation, Interpretation, information geometrical models, Internet of Things

Date début : 01.09.2021

---

1 Contexte de la thèse

L'apprentissage est essentiellement guidé par les données. Pourtant, les connaissances a priori, les invariants, la structure et les connaissances du domaine doivent apporter plus d'informations que les exemples d'apprentissage eux-mêmes. Dans certaines situations très peu d'exemples doivent être utiles pour apprendre. Ces principes doivent guider l'apprentissage automatique pour définir un apprentissage intelligent dans lequel la force brute des données doit laisser une grande place à la connaissance a priori.
L'équation fondamentale de l'apprentissage (Vapnik) résume l'apprentissage au processus de minimisation du risque à choisir une fonction qui explique une situation à partir d'un échantillon de données. Pour une fonction f appartenant à un ensemble donné de fonctions $\mathcal{F}$, une fonction de perte $L(x, y)$ qui calcule la perte à choisir y quand on sait que la solution est x et une distribution de probabilité P(x,y) inconnue, mais estimée par un échantillon de n exemples (x_i, y_i) sous contrainte iid, le risque à choisir f pour résumer toutes les données est :

$$R(f) = \int L(y, f(x))dP(x, y)$$

La résolution de cette équation a consisté principalement en l'estimation de la densité de probabilité à partir des données en se basant sur quelques résultats dans les plus connus sont : la capacité h de $\mathcal{F}$, le problème est soluble si et seulement si la VC dimension de $\mathcal{F}$ est finie et la limite supérieure du risque réel par le risque empirique par une quantité dépendant de h avec une probabilité supérieure à $1 - \eta$. $R(f) \leq R^n_{emp}(f) + \sqrt{(O(\frac{h-ln\eta}{n})(O(\frac{h-ln\eta}{n}) + 4R^n_{emp}(f)}$. D'autres résultats théoriques portant notamment sur le nombre nécessaires d'exemples pour apprendre (Hussler, Blumer, Walpert) permettent de mieux comprendre ce cadre.

L'étude de la partie droite de l'équation a conduit au développement d'une "force brute" de l'apprentissage par une demande importante de la quantité des données. La vitesse de croissance de l'erreur réelle vers l'erreur empirique dépend à la dimension de Vapnik-Chervonenkis (SRM) et du nombre d'exemples disponible obéissants aux contraintes de distribution.
Il s'agit de la tendance forte actuelle de l'apprentissage automatique portée par le deep learning et boostée par les succès technologique et industriels incontestables. Ceci se fait au détriment

de l'intégration des connaissances a priori et des invariants du domaine qui peuvent apporter plus de connaissances pour construire des modèles d'apprentissage sur des bases sûres et d'utiliser les données plus pour réfuter les aprioris et changer de modèles ou simplement améliorer de manière incrémentale les modèles existants que pour reconstruire totalement les solutions à partir de données seules.

La réécriture triviale de l'équation fondamentale (Vapnik) permet de faire apparaitre un terme totalement indépendant des données et s'intéresse donc à la partie droite de l'équation. Ceci suggère que la structuration de l'espace d'hypothèses peut être au moins aussi important que la disponibilité des données.

$$R(f) = \int (y - f_{best}(x)^2 dP(x,y) + \int (f_{best}(x) - f(x))^2 dP(x,y) + \int (y - fbest(x_)(f_{best} - f(x))dP(x,y)$$

Dans beaucoup d'applications réelles, la première intégrale est choisie a priori ce qui impose des biais incontrôlables et donc la nécessité d'avoir un espace de fonctions initiale à VC-dimension élevée pour corriger ces biais nécessitant de ce fait plus d'exemples que nécessaire. Nous pensons qu'il existe beaucoup de situations dans lesquelles le premier terme domine les deux suivants imposant, de ce fait, un cadre dans lequel les exemples nécessaires sont peu ou pas importants pour apprendre. Beaucoup des lois qu'on connait résultent d'expériences de pensées et les exemples servent à les confirmer ou à les réfuter. Quand les exemples réfutent la loi, ils servent à l'améliorer. Ça doit être aussi le cas en apprentissage. De plus, cette caractérisation hors de données suggère à la fois l'existence d'invariants exprimés par des concepts abstraits qui structurent le besoin des données et aussi une structuration explicite de l'espace d'hypothèses donnant de meilleures possibilités pour l'explication et l'interprétation des modèles d'apprentissage.

Un cadre simplifié dans lequel ce problème peut être étudié est l'internet des objets. Ceci pour plusieurs raisons parmi lesquelles : le fait que les objets soient totalement artificiels et contrôlables donc permettant un paramétrage des quantités, fréquences et autres hyperparamètres liés à la production des données, à la topologie et aux lois qui régissent les échanges. Le deuxième est qu'il est possible de choisir des structures appropriées pour l'analyse de certains phénomènes et tenter de ce fait de vérifier la validité de structures particulières de l'espace d'hypothèse (invariants, lois simples) nécessitant peu de données pour apprendre.

Les derniers travaux théoriques de Vapnik [12, 13, 11] sur cette question a donné naissance à une théorie dans laquelle le paradigme reflète totalement le travail que nous faisant avec une approche buttom-up en analysant les critères nécessaires (biais, prédicats, invariants, etc.) devant être intégrés dans un système d'apprentissage pour minimiser le nombre d'exemples nécessaires pour apprendre et pour accélérer la convergence ou simplement pour le placer dans un système proche de la connaissance a priori [5, 8, 9, 4, 10, 7]. Ce travail préliminaire nécessaire nécessite à la fois la proposition d'une architecture générale de métamodélisation et une formalisation mathématique du processus général de l'apprentissage dans des contextes concrets permettant de décrire dans un seul modèle les connaissances a priori, les invariants, l'explication, l'interprétation et la place des données dans le processus. Ce modèle de mataapprentissage doit permettre également de décrire un processus incrémental de l'apprentissage en terme d'évolution quantifiable de la qualité globale du modèle et son mode d'évaluation et non pas uniquement son adaptation en fonction des nouvelles données en entrée.

2 Défis scientifiques

Le défi de cette thèse est d'étendre au métalearning les derniers travaux de Vapnik sur la théorie complète de l'apprentissage dans laquelle il montre que la recherche des prédicats est un élément fondamental de l'apprentissage intelligent et l'appliquer au cadre de l'internet des objets dans lequel les caractéristiques sont : la coexistence de plusieurs apprenants, l'utilisation des mêmes sources de données pour apprendre des concepts antagonistes selon les points de vue et les objectifs, la structuration forte de certains aspects du contexte comme les protocoles de communication, la topologie de déploiement, le type de messages échangés, les capacités des divers composants. S'ajoute à cela, la manipulation de données essentiellement constituées de séquences temporelles et un changement continu du contexte et aussi la nécessité d'interpréter et d'expliquer les théories apprises par ces systèmes. Plus concrètement, parmi les défis scientifiques on notera une première focalisation sur la convergence faible ou la convergence en loi des propriétés plutôt que sur la minimisation du risque empirique qui se passe dans l'espace des fonctions des carrées intégrales et porte sur la convergence forte. Ce cadre pose plusieurs questions concrètes à aborder dans cette thèse la principale est d'intégrer la convergence faible dans le cas notamment des espaces de Hilbert à noyaux reproduisant pour des configurations simples. Ensuite, il convient d'exprimer certains invariants génériques qui nécessiteraient peu d'exemples (ou prédicats topologiques, géométriques, de structures, etc.) avec les outils appropriés. Les groupes de Lie offrent un cadre puissant pour l'expression des invariants, comme les invariants de transformation ou d'indépendance des systèmes de coordonnées [6, 1, 2], la préservation des invariances via les transformations, l'utilisation des symétries et pour les cas des séries temporelles, omniprésentes dans le cas de l'internet des objets, ils capturent naturellement la dynamique temporelle. Les modèles géométriques [1, 3] appliqués à des configurations spécifiques de l'internet des objets ainsi que des éléments de la théorie des représentations comme le théorème de Riesz [14, 15] pour mieux décrire au niveau structurel la notion d'interprétation et d'explication permettraient d'apporter des solutions liées à la confiance qui peut être accordée, par exemple, aux systèmes d'internet des objets devant prendre des décisions ayant des conséquences importantes sur l'environnement.

Références

[1] Frédéric Barbaresco. Higher order geometric theory of information and heat based on poly-symplectic geometry of souriau lie groups thermodynamics and their contextures : The bedrock for lie group machine learning. *Entropy*, 20(11) :840, 2018.

[2] Frédéric Barbaresco and François Gay-Balmaz. Lie group cohomology and (multi)symplectic integrators : New geometric tools for lie group machine learning based on souriau geometric statistical mechanics. *Entropy*, 22(5) :498, 2020.

[3] Adva Wolf Albert Gu Beliz Gunel Fred Sala, Ines Chami and Chris Ré. Into the wild : Machine learning in non-euclidean spaces "https ://dawn.cs.stanford.edu/2019/10/10/noneuclidean/". 10.

[4] Massinissa Hamidi and Aomar Osmani. Description of structural biases and associated data in sensor-rich environments. *CoRR*, abs/2104.04885, 2021.

[5] Massinissa Hamidi and Aomar Osmani. Domain models for data sources integration in HAR. *Neurocomputing*, 444 :244–259, 2021.

[6] Mei Lu and Fanzhang Li. Survey on lie group machine learning. *Big Data Min. Anal.*, 3(4) :235–258, 2020.

[7] Aomar Osmani and Massinissa Hamidi. Bayesian optimization of neural architectures for human activity recognition. In Nobuo Kawaguchi, Nobuhiko Nishio, Daniel Roggen, Sozo Inoue, Susanna Pirttikangas, and Kristof Van Laerhoven, editors, *Human Activity Sensing - Corpus and Applications*, pages 171–195. Springer, 2019.

[8] Aomar Osmani, Massinissa Hamidi, and Pegah Alizadeh. Augmented experiments in material engineering using machine learning.

[9] Aomar Osmani, Massinissa Hamidi, and Pegah Alizadeh. Hierarchical learning of dependent concepts for human activity recognition. In Kamal Karlapalem, Hong Cheng, Naren Ramakrishnan, R. K. Agrawal, P. Krishna Reddy, Jaideep Srivastava, and Tanmoy Chakraborty, editors, *Advances in Knowledge Discovery and Data Mining - 25th Pacific-Asia Conference, PAKDD 2021, Virtual Event, May 11-14, 2021, Proceedings, Part II*, volume 12713 of *Lecture Notes in Computer Science*, pages 79–92. Springer, 2021.

[10] Aomar Osmani, Massinissa Hamidi, and Salah Bouhouche. Monitoring of a dynamic system based on autoencoders. In Sarit Kraus, editor, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 1836–1843. ijcai.org, 2019.

[11] Vladimir Vapnik and Rauf Izmailov. Rethinking statistical learning theory : learning using statistical invariants. *Mach. Learn.*, 108(3) :381–423, 2019.

[12] Vladimir Vapnik and Rauf Izmailov. Complete statistical theory of learning : learning using statistical invariants. In Alexander Gammerman, Vladimir Vovk, Zhiyuan Luo, Evgueni N. Smirnov, Giovanni Cherubin, and Marco Christini, editors, *Conformal and Probabilistic Prediction and Applications, COPA 2020, 9-11 September 2020, Virtual Event, Verona, Italy*, volume 128 of *Proceedings of Machine Learning Research*, pages 4–40. PMLR, 2020.

[13] Vladimir Naumovich Vapnik. Complete statistical theory of learning. *Autom. Remote. Control.*, 80(11) :1949–1975, 2019.

[14] Whitney K. Newey Victor Chernozhukov† and Rahul Singhz. De-biased machine learning of global and local parameters using regularized riesz representers. *arXiv*.

[15] Haizhang Zhang, Yuesheng Xu, and Jun Zhang. Reproducing kernel banach spaces for machine learning. *J. Mach. Learn. Res.*, 10 :2741–2775, 2009.