

ENTREPRISE AUGMENTÉE

IA

Lev Kiwi, Assia Garbinato, Dominique Genoud, Giorgio Pauletto, Grégoire Ramuz



Par prédictibilité on entend la capacité des algorithmes à extraire à partir des données l'information déterminante pour le succès de la tâche. En d'autres termes, on imagine que les tâches les plus facilement prédictibles et ayant un coût d'erreur faible sont les premières à se faire automatiser. Par exemple, la maintenance prédictive est souvent un problème de « Machine Learning » ayant une relative-ment bonne prédictibilité. En revanche, lorsque le coût d'erreur est élevé l'automatisation est risquée. Pendant longtemps, ceci a été le cas pour la voiture autonome. Cependant, depuis les progrès dans le domaine, utilisant des réseaux de neurones, la prédictibilité a augmenté significativement rendant le risque d'automatisation plus faible.

Dans un premier temps, l'automatisation de tâches intelligentes va se développer dans les étapes des processus d'entreprise à faible valeur ajoutée. Le but étant de libérer l'homme de tâches peu intéressantes. Les tâches ayant un coût d'erreur plus important ne vont probablement pas être totalement automatisées. Ces tâches bénéficieront d'un support algorithmique dans l'optique d'une aide à la décision. On parle alors d'**intelligence augmentée**. On imagine des partenariats homme-machine afin de bénéficier du meilleur des deux mondes. Ceci est d'autant plus intéressant que la machine ne fait pas souvent les mêmes erreurs que les humains. Les algorithmes auto-apprenants n'implémentent pas de raisonnement logique et se basent principalement sur des statistiques. Ils n'utilisent aucunement les informations contextuelles autour de la donnée pour parvenir à la prédiction. C'est en combinant l'intelligence artificielle avec les compétences cognitives humaines que l'on s'attend à obtenir en général de meilleurs résultats.

Ce partenariat homme-machine va nous inviter à donner du feedback à la machine. Actuellement, ceci n'est fait que de façon anecdotique dans certaines applications. Cependant, afin que le métier adopte ces technologies, il

sera primordial d'afficher une certaine transparence dans le fonctionnement de ces algorithmes. Le métier aura besoin de comprendre au moins d'une façon grossière les mécanismes sous-jacents. L'utilisateur aura ainsi son mot à dire sur les biais potentiels que l'algorithme peut identifier statistiquement. Cette compréhension pourra même être poussée jusqu'à l'interaction directe du métier avec les paramètres des modèles statistiques. Il est tout à fait imaginable qu'à l'avenir on utilise principalement des intelligences artificielles de type *boîtes noires* comme c'est encore le cas. Cela se résume à perdre le contrôle sur la machine et personne ne souhaite cela.

En conséquence, le partenariat homme-machine va ainsi donner un nouveau rôle à l'homme, celui de l'enseignant de la machine. Dans le contexte déterministe de la programmation, l'unique enseignant est le développeur. Dans un contexte probabiliste, l'utilisateur peut aussi devenir l'enseignant. Évidemment, il est encore délicat de laisser un algorithme auto-apprenant de façon autonome dans la nature. Microsoft en a fait les frais avec son chatbot Tay lorsque ce dernier s'est mis à tenir des propos racistes. En effet, l'intelligence artificielle s'est fait « hacker » très simplement car elle adapte son langage en fonction des tweets des utilisateurs. Il a suffi que des hackers postent des propos racistes en masse sur Twitter pour que l'algorithme les intègre dans sa conversation.

Les algorithmes de « Machine Learning » supervisés et non supervisés font l'objet aujourd'hui d'une grande maturité. En revanche, il n'en est pas de même des algorithmes d'apprentissage par renforcement. A notre avis, ils sont probablement sous-utilisés. Ces algorithmes capables d'apprendre et de s'adapter à chaque instant vont probablement faire l'objet d'intenses recherches ces prochaines années. Le succès d'AlphaGo Zero en 2017 nous montre qu'une approche par renforcement permet d'aller bien au-delà des limites de l'apprentissage supervisé.

L'apprentissage par renforcement nous permettra aussi de créer des interfaces auto-adaptatives dans les applications. A priori, toute application à l'avenir sera capable d'apprendre l'utilisation qu'on en fait. Elle sera ainsi capable d'adapter son interface afin de rendre son utilisation plus facile. On le voit déjà avec certains sites web, principalement de news, qui ajustent le contenu en fonction de notre comportement sur le site en question. On imagine que les applications de demain ne se contenteront pas d'adapter leurs contenus à l'utilisateur, mais aussi leurs fonctionnalités.

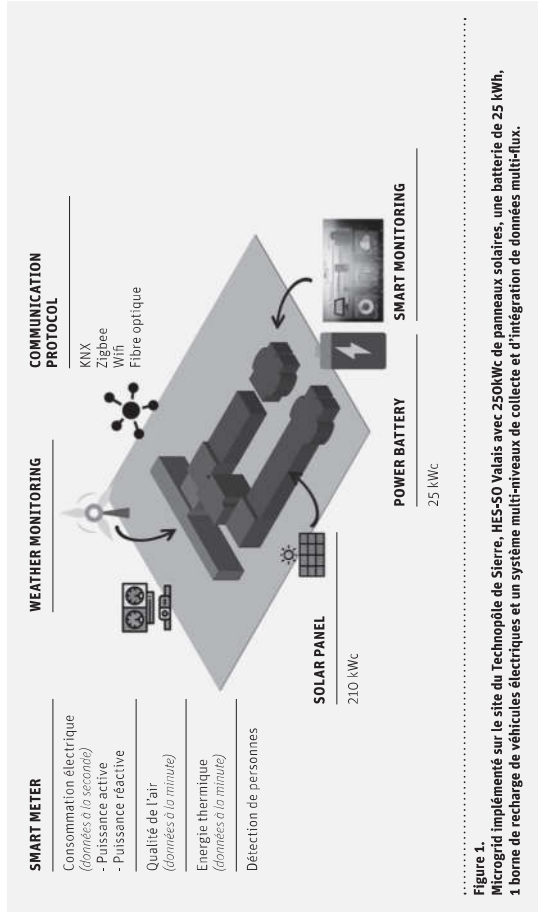
Au bout du compte, l'apprentissage par renforcement nous donnera accès à un nouveau paradigme de programmation dans le « Machine Learning », celui de la conception d'algorithme automatisée, ou *automated algorithm design*. Cette couche d'abstraction est composée de méta-algorithmes de « Machine Learning ». En d'autres termes, des algorithmes de « Machine Learning » qui conçoivent des algorithmes de « Machine Learning ». Actuellement, leur processus de conception nécessite un travail fastidieux d'affinage de paramètres. Avec le « Deep Learning », le nombre de paramètres à ajuster a simplement explosé, rendant la tâche épuisante et inintéressante. Ainsi la conception d'algorithme automatisée est un contexte dans lequel un algorithme optimise un algorithme de « Machine Learning » pour une tâche spécifique. Google s'est déjà lancé dans ce créneau et contribue de façon significative au projet de recherche « Machine Learning for Automated Algorithm Design ».

4 UNE ÉTUDE DE CAS : AIDE DU PILOTAGE ÉNERGÉTIQUE DÉCENTRALISÉ PAR PRÉDICTION

Cette étude de cas est ici développée sous la plume de Dominique Genoud, David Wannier, Mariam Barque, Luc Dufour, HES-SO/Valais-IG.

Introduction
La transition énergétique passe par une maîtrise de l'injection dans le réseau électrique de la production d'énergie renouvelable d'origine éolienne et photovoltaïque. D'autant plus que la majorité des installations solaires déployées en France et en Suisse sont de faibles puissances, raccordées au réseau basse tension, lieu où l'information sur leur comportement à court terme manque souvent. Les modèles prédictifs historiques sont centralisés et paramétriques; ils ont du mal à appréhender les variations brutales de productions et de consommations, et un maillage plus fin devient donc utile pour comprendre les scénarios d'usage dans les bâtiments.

L'énergie renouvelable a permis d'amorcer un changement de paradigme de notre système énergétique, mais il introduit des contraintes supplémentaires sur le réseau électrique. Les énergies renouvelables sont par définition des énergies dites de « flux », par opposition aux énergies fossiles et nucléaires qui reposent sur des « stocks ». Leur disponibilité est ainsi soumise aux aléas météorologiques plus ou moins prévisibles. Comme la contrainte majeure de la stabilité d'un réseau électrique est de maintenir l'équilibre constant entre la production et la consommation, les gestionnaires de réseau veillent donc à tout moment à satisfaire exactement la consommation, ni plus, ni moins. Pour parer à l'intermittence des énergies renouvelables, les distributeurs doivent adapter la quantité d'énergie apportée par d'autres moyens de production plus flexibles comme l'hydraulique, les centrales à énergie fossile ou utiliser des batteries de stockage. La prédiction de consommation et de production leur permet ainsi d'équilibrer toutes ces sources d'énergie.



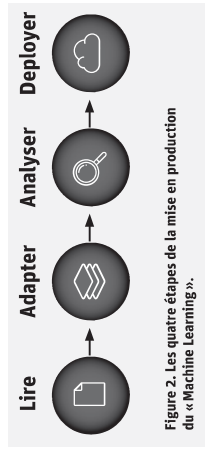
Méthodes de prédiction des flux énergétiques

Il existe des moyens de reproduire les grands flux énergétiques d'un bâtiment en utilisant par exemple des logiciels de simulation thermique détaillée ou de production solaire pour construire des simulations horaires. Cependant, il reste de fortes incertitudes sur les prédictions locales et précises dans le temps avec parfois de grandes différences entre les simulations et la réalité. Pour tenter de diminuer ces erreurs, nous avons utilisé une approche complémentaire basée sur le « Machine Learning », qui va apprendre à prédire le comportement du système à partir des données pour compléter de manière plus dynamique les prédictions à court terme. Cette approche à deux niveaux va nous permettre de construire des modèles de référence à partir desquels des études énergétiques

peuvent être réalisées et servir de modèles de prédictions de base quand le système n'a pas encore récolté de données. Puis dans un second temps, elle va nous permettre d'utiliser des méthodes d'apprentissage pour construire des modèles basés sur le passé qui caractérisent la prédiction pour prédire le futur. Comme les modèles de prédictions de consommation possibles sont très nombreux, les échelles de temps sont multiples, (au 1/4 d'heure, à l'heure, à la journée, au mois ou même à l'année) les possibilités d'apprentissage sont nombreuses et les modèles d'apprentissage utilisés doivent être généralisables. Ainsi, nous avons testé des méthodes paramétriques et non paramétriques, comme par exemple des réseaux de neurones ou des ensembles d'arbres de décision [LUC-2017].

Méthodologie de prédiction par « Machine Learning »

De manière générale, la prédiction par « Machine Learning » se déroule en quatre étapes, comme le montre la figure 2.



Etape 1 : Sources de données

Cette étape orchestre les différentes sources de données disponibles pour qu'elles représentent une information cohérente et utilisable. Dans le cas des mesures énergétiques pour le bâtiment, il s'agit de l'historique des mesures de la production, de la consommation des différentes entreprises du site du technopôle (voir figure 1), d'informations météo mesurées localement ainsi que des prévisions météo fournies par MétéoSuisse.

Etape 2 : Transformation des données

Cette deuxième étape est souvent la plus délicate car le but est de nettoyer les données et d'enlever le maximum de sources d'erreurs possibles avant l'utilisation des algorithmes de modélisation (classification, prédiction) pour éviter qu'elles ne se répercutent dans les résultats (valeurs manquantes, heures de nuit pour le solaire, périodes de maintenance pour l'éolien, etc.).

Etape 3 : Apprentissage et validation de la prédiction

Cette étape va entraîner et valider les algorithmes à apprentissage qui vont créer un modèle de la réalité appris sur les données. Ce modèle va ensuite être utilisé pour prédire le comportement futur de la production ou de la

consommation du système énergétique. Pour ce faire, l'ensemble des données est habituellement séparé en deux : une partie des données servant à apprendre une modélisation et une autre partie à valider la qualité de l'apprentissage et ainsi choisir le type d'algorithme qui produit les meilleures performances. Dans le présent exemple, nous allons utiliser une prédiction soit de classes de valeurs de prédiction (classification) soit directement de valeurs numériques (voir le chapitre 1).

Etape 4 : Déploiement du « Machine Learning »

Cette étape permet d'utiliser les modèles appris lors de l'étape 3 avec des données nouvelles et de proposer des prédictions de manière régulière. Ces prédictions peuvent ensuite être visualisées pour permettre de prendre des décisions, ou alors réinjectées dans un système d'information pour contrôler par exemple la production ou décaler la consommation.

Comment obtenir des modèles de « Machine Learning » performants

Dans les résultats présentés ci-dessous, le vecteur d'entrée qui caractérise la consommation électrique globale regroupe les statistiques liées à la puissance et aux différentes variations au niveau du compteur global d'électricité. La source des données météorologiques est constituée des valeurs horaires réelles et prédites des stations géographiquement proches de notre cas d'étude. La temporalité de notre signal est gardée et délimitée par les variables minutes, jours, jours de la semaine et heures. Pour les résultats de la prédiction, chaque heure est labélisée par une classe de sortie qu'il s'agira de déterminer à partir de la modélisation. Afin de préparer correctement l'ensemble des données, des informations complémentaires (métadonnées), comme les heures d'arrivée et de départ du personnel ainsi que les jours travaillés, sont rajoutées et permettront d'affiner la qualité du jeu d'entraînement en filtrant des données qui n'apportent aucune information.

Dans le secteur tertiaire, les jours et les heures en dehors du planning de production sont enlevés de notre jeu de données d'entraînement et de test. Pour le restaurant, le week-end est enlevé et seules les heures entre 6 h et 17 h sont conservées. Pour notre étude, l'ensemble du jeu de données rassemble 4024 heures du 1^{er} septembre 2013 au 31 octobre 2014 (déduction faite des données non pertinentes comme expliqué ci-dessus). Finalement, 3772 heures sont utilisées pour entraîner le modèle et 252 heures représentent notre jeu de test. Le tableau 1 montre que les performances sont excellentes et indiquent un coefficient de détermination R^2 élevé (voir encadré).

Après une première modélisation qui a utilisé des réseaux de neurones de plusieurs sortes ainsi que des forêts d'arbres de décision de plusieurs types, nous avons obtenu les meilleurs résultats avec un ensemble d'arbres de décision de type particulier : le Gradient Boosted Tree (voir glossaire).

R^2

En statistique, le coefficient de détermination, noté R^2 permet de déterminer à quel point l'équation de régression est adaptée pour décrire la distribution des points.

Si le R^2 est nul, cela signifie que l'équation de prédiction détermine 0% de la distribution des points. Dans ce cas, le modèle mathématique utilisé n'explique absolument pas la distribution des points.

Si le R^2 vaut 1, cela signifie que l'équation de prédiction est capable de déterminer 100% de la distribution des points. Dans ce cas, le modèle mathématique utilisé détermine parfaitement le phénomène à prédire.

En définitive, plus le R^2 est proche de 1, plus le modèle prédictif est proche de la réalité.

Afin de compléter les mesures de précision de la prédiction on ajoute habituellement d'autres mesures statistiques qui permettent de se faire une idée plus précise des erreurs et de leur forme comme le MAE (Mean Absolute Error), le MSE et RMSE (Mean Squared Error et sa racine), le MSD (Mean Square Deviation) (voir le glossaire).

Gérer le « Machine Learning » dans le temps : le réapprentissage des modèles

Le réapprentissage appliqué consiste à utiliser de nouvelles données pour compléter les modèles existants. Pour une prédiction du jour suivant par exemple, le nouvel ensemble de données utilisées pour prédire le jour $J+1$ est obtenu par la concaténation des données utilisées pour prédire le jour J et des dernières mesures de production et de météo récoltées le jour J . L'ensemble d'apprentissage s'agrandit donc afin de prendre en compte les informations accessibles au plus proche de la période à prédire. Dans certaines approches, la phase d'apprentissage est réalisée une fois et le modèle établi permet de calculer la prédiction pour tous les jours ou les heures qui suivent. L'approche de réapprentissage suivie ici permet de commencer la prédiction par « Machine Learning » avec une quantité de données limitée, mais aussi d'adapter la prévision en fonction des valeurs les plus proches de l'instant à prédire, comme c'est le cas pour les modèles autorégressifs. Cependant, un suivi et une évaluation de ces modèles avec des données historiques et une analyse des erreurs sont importants pour savoir quand stopper le réapprentissage.

Exemple du solaire et de l'éolien

Nous avons utilisé cette méthode pour la prédiction de production solaire, en utilisant des données de prévision météorologique et des données de mesures locales afin de diminuer l'erreur de prévision à un moment et à un endroit donné. En effet, les prévisions météo sont calculées pour une résolution spatiale qui peut aller de 1 km à 30 km. Le fait d'utiliser les données réelles mesurées au niveau de la centrale solaire permet en quelque sorte d'affiner la résolution spatiale des prévisions. Ce réapprentissage utilisant les erreurs météo permet de gagner 10 % de précision dans le cas de la prédiction de production solaire du jour suivant. Une approche similaire a été appliquée à la prédiction éolienne avec une amélioration de 6 % de précision pour le jour suivant.

Optimiser les modèles de « Machine Learning » : La réduction du nombre de variables

Une des grandes questions du « Machine Learning » est l'usage adéquat des variables d'entrées, leurs corrélations, leurs variabilités et l'impact qu'elles ont sur la modélisation. Ce sont des questions parfois difficiles à résoudre. De nombreuses méthodes sont à disposition pour tenter de réduire et rendre plus cohérent l'espace des variables d'entrées. Dans le cas présenté ici, comme nous avons utilisé des forêts d'arbres de décision, nous connaissons de manière explicite la statistique des variables d'entrées utilisées par nos arbres. Nous sélectionnons ensuite les variables par ordre décroissant de fréquence d'apparition, et le calcul prédictif sera réalisé en faisant diminuer le nombre de variables basées selon leurs importances dans la construction des arbres.

Le **Tableau 1** montre les résultats obtenus sur un grand bâtiment possédant de nombreuses variables. Cette approche se révèle intéressante car elle permet de réduire fortement le nombre de capteurs et de mesures à effectuer en maîtrisant la qualité du modèle. Dans certaines applications, le nombre de variables d'entrée peut être réduit considérablement sans baisse de performance de la prédiction. On observe que même en réduisant le nombre de variables d'entrées de plus de 4000 à 50 le R^2 reste encore très important.

Nombre de variables	MAE	MSE	RMSE	MSD	R^2
4160	1.488	4.494	2.12	0.124	0.946
150	1.882	7.583	2.754	-0.053	0.908
50	2.002	8.283	2.878	-0.343	0.899
15	2.22	10.767	3.281	-0.109	0.869

Tableau 1 : Résultats statistiques en sortie du Gradient Boosted Tree en fonction du nombre de variables pour la prédiction de la consommation électrique globale pour le secteur tertiaire

Conclusion

Les travaux de l'Institut Informatique de Gestion de la HES-SO Valais Wallis (IIG) dans le domaine de la prédiction d'énergie renouvelable ont montré l'efficacité et la robustesse du « Machine Learning » dans les approches de prédiction de flux énergétiques décentralisés (quartiers, bâtiments industriels). Une plateforme prédictive de la production du parc solaire de l'ESR (Energie Sion Région), distributeur régional d'électricité, a été mise en place en 2015 en collaboration avec l'Energy center. La prédiction de la production du parc éolien valaisan exploitée par SEIC-Télédis SA a également été réalisée et testée sur la plateforme web de visualisation « Portail des Energies Renouvelables du Valais ». La méthodologie utilisée a fait l'objet d'une publication. De nombreux Best Papers ont été remportés dans les domaines de la prédiction de l'énergie renouvelable et de la consommation d'électricité dans les ménages et les industries.

Auteurs : Dominique Genoud, David Wannier, Mariam Barque, Luc Dufour, HES-SO//Valais-IIG