

L'intelligence artificielle au service des réseaux

Le projet Grid Data Digger | Comment garantir un fonctionnement optimal et sécurisé des réseaux de distribution d'électricité dans un contexte d'évolution constante ? La plateforme « Grid Data Digger » utilise des méthodes d'analyse de données massives (big data) et d'apprentissage automatique (machine learning) pour aider les gestionnaires de réseaux à relever ce défi.

M. BOZORG, N. FATEMI, C. PENA, O. MOUSAVI, M. CARPITA

Dans le contexte de la transition énergétique, les réseaux de distribution d'électricité doivent faire face à des changements majeurs. Ces derniers se manifestent dans la production d'énergie avec l'introduction croissante d'énergie renouvelable, distribuée et intermittente (à l'instar des systèmes photovoltaïques), mais aussi dans les habitudes de consommation énergétique : l'adoption des véhicules électriques (+144 % d'immatriculations entre 2018 et 2019 en Suisse [1]), notamment, change rapidement le profil de charge des consommateurs et augmente donc sa variabilité.

Face à ces changements fondamentaux, les gestionnaires de réseau de distribution (GRD) doivent adapter les processus d'exploitation et de gestion de réseau pour intégrer ces nouvelles productions distribuées et consommations stochastiques dans les réseaux de moyenne et basse tension (MT/BT), et ce, sans compromettre la sécurité des réseaux ni la qualité de service.

Adapter les processus grâce aux données

Aujourd'hui, en Suisse, les GRD utilisent dans leurs réseaux de distribution MT et BT des appareils de mesure

et d'acquisition de données de réseau en temps réel tels que la solution GridEye de la société Depsys : cette solution « plug-and-play » analyse les réseaux de distribution grâce au « edge computing » (informatique en périphérie) et transfère les données de mesure à la plateforme en ligne pour une analyse en temps réel. Ces appareils de mesure installés sur des réseaux BT fournissent une grande quantité de données. Cependant, ils ne permettent aujourd'hui qu'une surveillance partielle car l'interprétation de ces données, et par conséquent la prise de décision, reste difficile.

Le projet Grid Data Digger (GDD), un projet soutenu par Innosuisse, a pour objectif d'interpréter et de valoriser ces données par le biais de processus automatisés, afin de fournir aux GRD toutes les informations et analyses nécessaires. Il leur permettra d'identifier les tendances majeures dans le réseau, de déceler des irrégularités, de prédire des dépassements de limites d'exploitation, d'analyser leur origine et de mener des actions de contrôle préventif afin d'éviter ces dépassements. La solution développée dans le contexte de ce projet comprend une plateforme « big data » gérant les flux de données massives grâce à des applications d'analyse de données descriptives, diagnostiques, prédictives et préventives dédiées à l'exploitation du réseau de distribution.

Le défi de la gestion des flux de données massives

La **plateforme big data** du projet GDD offre l'infrastructure comprenant tous les outils et les composants nécessaires pour l'ingestion, l'analyse intelligente et le stockage des quantités massives de données du réseau récoltées par des appareils GridEye (figure de titre). Sa conception est basée sur des études et expérimentations comparatives des technologies les plus récentes du domaine du big data, notamment sur les résultats de tests appliqués aux données (réelles et simulées). Ceci permet de mesurer divers paramètres de performance tels que la quantité des ressources requises, la qualité et la vitesse d'analyse, la capacité de montée en charge, etc.

L'**ingestion des données massives** assure un acheminement robuste et en continu des données du réseau récoltées vers les outils d'analyse intelligente et le stockage des données. Le défi consiste à assurer la collecte et la distribution d'une quantité massive de données sans pertes, et ce, malgré les problèmes potentiels de connectivité du réseau ou des appareils de mesure.

Pour certaines applications, une analyse en temps réel est nécessaire afin de permettre une prise de décision dans les quelques secondes ou minutes suivant la récolte des données. Pour ce faire, une technologie appelée « traitement de flux de données » (stream processing) est utilisée. Un exemple typique serait la prédiction, grâce à

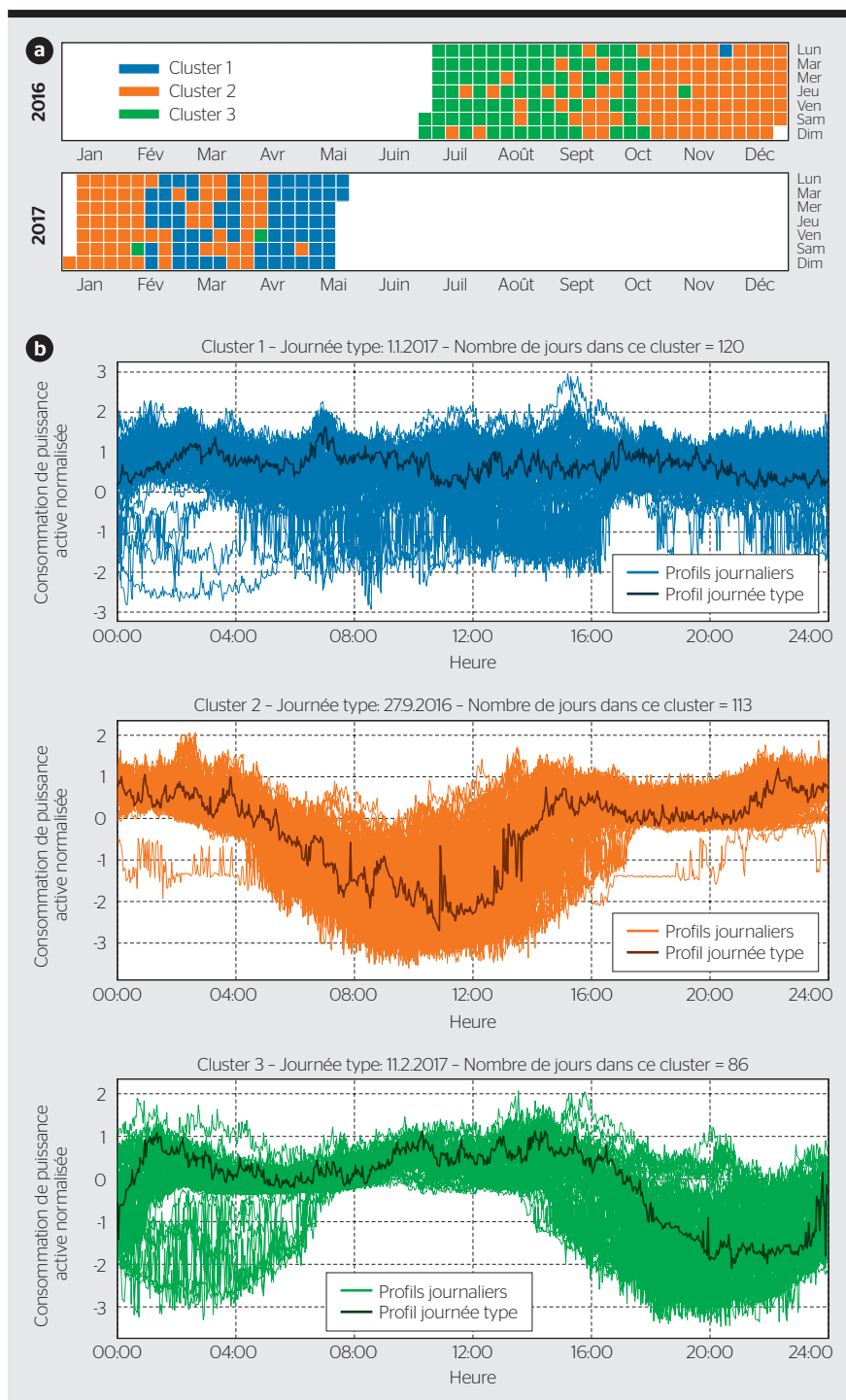


Figure 1 Regroupement des données dans trois clusters, chaque cluster représentant une journée type (representative day) (a), et profils normalisés de puissance de consommation pour trois journées types (b).

l'utilisation d'un flux de données en temps réel, de la surcharge des transformateurs ou des câbles de réseaux quelques heures avant que celle-ci n'intervienne.

L'**analyse intelligente des données** rassemble les techniques et les algorithmes d'apprentissage automatique (machine learning) spécifique-

ment conçus et adaptés aux analyses requises par la plateforme GDD. Les modèles d'apprentissage sont développés et validés en phases itératives à partir des données mesurées sur les réseaux électriques. Afin d'assurer la scalabilité de la plateforme big data, les algorithmes sont ensuite adaptés à une exécution sous Apache Spark MLLib,



Figure 2 Exemple de caractéristiques significatives (minimum/maximum annuels de la tension et de la puissance normalisés) pour la classification des 729 nœuds d'un réseau électrique simulé en fonction de la capacité photovoltaïque installée. Chaque point représente les séries temporelles des tensions et puissances de consommation normalisées d'un nœud du réseau pendant un an.

une librairie de machine learning distribuée, puis mis à l'épreuve lors de tests de montée en charge grâce à un simulateur de réseaux électriques.

Enfin, le **stockage des données** permet la persistance à moyen et à long terme des mesures récoltées. Les données sont tout d'abord mises à disposition des algorithmes d'analyse, par exemple celui de la classification de niveau d'énergie photovoltaïque produit dans chaque quartier. Par la suite, un historique des données sur plusieurs années permettra d'affiner ces modèles de machine learning avec l'ajout des nouvelles données. La plateforme GDD met à disposition des moyens de stockage qui optimisent la recherche et l'agrégation de ces données temporelles. GDD offre un stockage distribué des données sur plusieurs machines reliées dans un cluster qui peut être,

selon les besoins ou les souhaits, géré sur place (on-premises) ou dans le cloud.

Vers une gestion adaptée et prédictive du réseau

L'intelligence artificielle (IA), ou plus exactement le « machine learning », donne du sens et de la valeur aux immenses quantités de données générées, récoltées et stockées par les appareils de mesure existants sur les réseaux. En puisant toujours dans l'information des signaux de courant, de tension et de puissance, l'IA est utilisée par exemple pour identifier différents scénarios d'exploitation du réseau, pour comprendre les types d'utilisation et les profils de charge, ou pour prédire de potentiels dépassements de la limite d'exploitation du réseau. Pour cela, une large palette de méthodes est appliquée,

allant de l'apprentissage non supervisé des algorithmes de regroupement (clustering), à la classification et la prédiction des séries temporelles par des algorithmes d'apprentissage supervisé.

Décoder les motifs récurrents grâce au clustering

Si, par la pratique, chaque GRD sait que l'exploitation du réseau diffère d'une saison à une autre et d'un jour ouvrable à un jour férié, il est difficile d'établir aussi intuitivement le nombre de journées types existant réellement. Utiliser des algorithmes de regroupement permet de découvrir des groupes de données, des journées dans notre cas, présentant des caractéristiques similaires entre elles, mais différentes de celles des autres groupes.

Différents algorithmes ont été examinés, par exemple des algorithmes de regroupement classiques (*k*-means ou *k*-medoids) ainsi que des algorithmes beaucoup plus récents et en cours de développement tels que Matrix Profiles. L'exemple de la **figure 1** permet de distinguer clairement trois groupes, correspondant dans ce cas aux saisons : été (vert), automne-hiver (orange) et printemps (bleu), ainsi que des journées atypiques.

Grâce à cette méthode, il est possible de traiter en quelques secondes les données de plusieurs années, y découvrir des motifs récurrents, évaluer s'ils sont cycliques sur le long terme et repérer rapidement des comportements qui se distinguent.

Identifier les profils grâce à l'apprentissage supervisé

Les analyses de données permettent également de comprendre les types et profils de consommation et leur utilisation afin de mieux prévoir la demande d'énergie et donc d'améliorer la planification de la gestion du réseau.

Pour identifier ces profils, les données temporelles sont exploitées afin de construire, à l'aide d'algorithmes d'apprentissage supervisés, des classificateurs du type d'utilisation (commerciale, résidentielle ou industrielle), de la capacité photovoltaïque installée et de la capacité des bornes de recharge pour véhicules électriques (zéro, basse, moyenne ou élevée). Ces classificateurs sont entraînés en utilisant des données simulées, produites par des modèles propriétaires développés dans le cadre du projet.

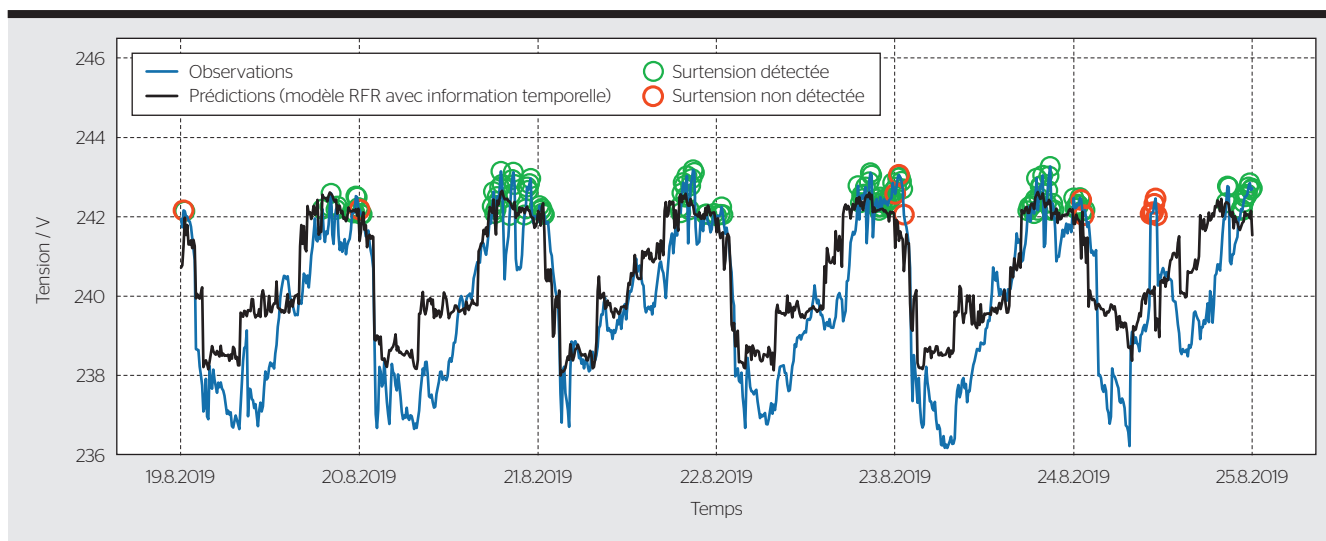


Figure 3 Exemple de résultats de prédiction des surtensions sur un nœud de réseau BT (hypothèse : seuil de surtension égal à 242 V). L'horizon de prédiction est de 4 heures.

La classification des séries temporelles diffère du problème de classification habituel dans le sens où les données sont ordonnées dans le temps. Pour aborder ce problème, chaque série est représentée par un ensemble de caractéristiques significatives (un exemple se trouve à la **figure 2**). Ces caractéristiques sont ensuite utilisées pour construire des modèles prédictifs permettant de classer les séries temporelles dans les différentes catégories. Pour ce faire, plusieurs méthodes d'apprentissage automatique ont été appli-

quées de façon à comparer leurs performances, parmi lesquelles : la méthode des k plus proches voisins (k -NN), l'apprentissage statistique (naïve Bayes), les machines à vecteurs de support (SVM) ou les forêts d'arbres décisionnels (random forests). Des modèles prédictifs sont ainsi obtenus, permettant par exemple de déterminer si le type de consommation correspond à une zone résidentielle ou commerciale avec plus de 90 % de précision, ceci sur la base des caractéristiques de consommation agrégées par

heure de la journée et modélisé par une série de règles fournies par un arbre de décision.

Anticiper les surcharges grâce aux algorithmes de prédiction

Le réseau électrique est soumis à des perturbations, telles que des surcharges, qui peuvent l'amener à un état indésirable, c'est à dire à dépasser les limites acceptables d'exploitation. Les algorithmes de prédiction vont permettre d'anticiper les flux de puissance et l'état du réseau (les tensions), et ainsi

IN KÜRZE

Die künstliche Intelligenz im Dienst der Netze

Das Projekt Grid Data Digger

Die wachsende Einspeisung von schwankenden erneuerbaren Energien sowie die Veränderungen im Verbrauchsverhalten insbesondere durch die Einführung der Elektromobilität stellen eine Herausforderung für Verteilnetzbetreiber (VNB) dar. Wie lässt sich ein optimaler und zuverlässiger Betrieb der Verteilnetze im Kontext dieser dynamischeren Situation gewährleisten? Heute setzen die VNB Mess- und Datenerfassungsgeräte ein, die in Echtzeit arbeiten. Obwohl Letztere eine riesige Datenmenge liefern, erlauben sie nur eine partielle Überwachung, weil die Interpretation dieser Daten anspruchsvoll bleibt.

An der HEIG-VD (Haute école d'ingénierie et de gestion du canton de Vaud) wird zurzeit die Plattform «Grid Data Digger» entwickelt, die Big-Data-Analysen und Machine-Learning-Methoden einsetzt, um u. a. Unregelmässigkeiten im Netz zu erkennen und Überschreitungen von Spannungsgrenzen zu prognostizieren. In diesem Projekt werden zahl-

reiche Methoden eingesetzt, vom unüberwachten Lernen mit Clustering-Algorithmen bis hin zur Klassifizierung und Prognose von zeitlichen Serien durch «Überwachtes Lernen»-Algorithmen. Die Entwicklung der Plattform basiert auf Studien und Experimenten, bei denen die neusten Big-Data-Technologien verglichen werden, besonders auf Testresultaten, die mit realen und simulierten Daten durchgeführt wurden. Dies ermöglicht es, diverse Leistungsparameter zu messen wie beispielsweise die Menge der benötigten Ressourcen, die Qualität und Geschwindigkeit der Analyse und die Skalierbarkeit.

Die Plattform hat beispielsweise bereits eine Genauigkeit von 82 % bei der Prognose von Über- und Unterspannungen in einem Prognosezeitraum von vier Stunden erreicht. Weitere Methoden, insbesondere Deep Learning Methoden, werden untersucht mit dem Ziel, die Prognosegenauigkeit noch zu steigern.

CHE

de prédire l'occurrence d'un tel état indésirable quelques heures avant l'événement lui-même, de sorte que le GRD ait suffisamment de temps pour prendre des mesures correctives et préventives.

Pour entraîner ces modèles de prédiction, des données de tension, de courant et de puissance ont été utilisées. Ces mesures correspondent à une année d'exploitation et les valeurs sont moyennées toutes les 10 minutes. Ces données ont ensuite été enrichies avec des informations telles que l'heure, le mois, le jour de la semaine et les jours fériés.

Deux approches ont été testées. Dans la première, les séries temporelles sont simplement classifiées comme pouvant mener ou non à un état indésirable dans les heures suivantes. La seconde prédit d'abord les valeurs probables de la tension pour les prochaines heures et détecte ensuite, dans ces valeurs prédites, d'éventuels dépassements des limites d'exploitation. Les meilleurs résultats ont été obtenus avec cette seconde approche qui, en utilisant la méthode de régression par forêts aléatoires (RFR, random forest regressor), permet d'atteindre un taux de précision de 82% (et un taux de rappel de 75%) dans la prédiction des surtensions et sous-tensions dans un horizon de prédiction de 4 heures. À titre d'exemple de résultats obtenus, la **figure 3** montre les surtensions détectées correctement

par l'algorithme, ainsi que celles qui ne sont pas détectées. Cette performance devient encore meilleure en utilisant des horizons de prédiction de 2 heures (83% et 80%) ou d'une heure (86% et 78%).

Par la suite, d'autres méthodes vont être explorées dans le but d'améliorer encore ces performances, notamment des méthodes issues de l'apprentissage profond (deep learning) telles que les réseaux de neurones convolutifs unidimensionnels (1D-CNN), les réseaux neuronaux récurrents (RNN) et des variantes comme la mémoire à long-courte terme (LSTM).

Une validation à grande échelle

Le projet Grid Data Digger a déjà permis des avancées majeures vers une gestion encore plus intelligente, prédictive et préventive des réseaux de distribution d'électricité.

Les résultats du projet sont validés à petite échelle dans le laboratoire Réseaux intelligents ReIne [2-3] de la HEIG-VD. Les prochaines étapes se concentrent désormais, d'une part, sur la validation et l'application de ces résultats à grande échelle et, d'autre part, sur le développement et la mise en œuvre des nouveaux algorithmes enrichissant les possibilités de détection d'événements, d'irrégularités et d'alertes pour éviter les dépassements des limites du réseau. L'objectif ultime consiste à fournir aux GRD toutes les

informations et analyses nécessaires au fonctionnement sécurisé et optimal des réseaux de distribution.

Références

- [1] Office fédéral de la statistique, communiqué de presse, « Véhicules routiers en 2019: parc et nouvelles mises en circulation », 31 janvier 2020. www.bfs.admin.ch/bfs/fr/home.assetdetail.11407549.html.
- [2] M. Carpita, J. Affolter, M. Bozorg, D. Houmard and S. Wasterlain, « ReIne, a flexible laboratory for emulating and testing the Distribution grid », 2019 21st European Conference on Power Electronics and Applications (EPE '19 ECCE Europe), Genova, Italy, 2019. ieeexplore.ieee.org/document/8915190.
- [3] M. Carpita, J.-F. Affolter, et M. Capezzali, « Relever les défis de la transition énergétique », Bulletin SEV/VSE 10/2018, pp. 40-43, 2018. bulletin.ch/fr/news-detail/relever-les-defis-de-la-transition-energetique.html

Auteurs

Prof. Dr Mokhtar Bozorg est professeur en énergie et réseaux électriques à la Haute école d'ingénierie et de gestion du canton de Vaud (HEIG-VD).
→ HEIG-VD, 1401 Yverdon-les-Bains
→ mokhtar.bozorg@heig-vd.ch

Prof. Dr Nastaran Fatemi est professeure en systèmes d'information et analyse de données big data à la HEIG-VD.
→ nastaran.fatemi@heig-vd.ch

Prof. Dr Carlos Andres Pena est professeur en apprentissage automatique et applications biomédicales à la HEIG-VD.
→ carlos.pena@heig-vd.ch

Dr Omid Mousavi est directeur R&D chez Depsys.
→ Depsys, 1070 Puidoux
→ omid.mousavi@depsys.com

Prof. Dr Mauro Carpita est professeur en électronique de puissance et directeur de l'Institut IESE à la HEIG-VD.
→ mauro.carpita@heig-vd.ch

Les auteurs désirent remercier Dr Mohammad Rayati, Marten Fesefeldt, Christopher Meier, Guillaume Hôchet, Oscar Rodriguez, Dr Shabnam Ataee et Julian Burella (HEIG-VD) pour leur contribution au projet, ainsi que Marianne Tamborini (Depsys) pour son soutien dans la rédaction de cet article.