



EDF R&D : PROPOSITION DE STAGE DE FIN D'ETUDES (6 MOIS)

Renforcement des méthodes de désagrégation des séries temporelles de consommation électrique par Active Learning

Le contexte :

Le département **SEQUOIA** (Services, Économie, Questions hUmaines, Outils innovants et IA) de la R&D d'EDF soutient la branche Commerce en proposant des approches innovantes permettant d'approfondir la connaissance client. Au sein de ce département, le groupe **E7C** est spécialisé dans l'analyse de données de consommation et la modélisation statistique, afin de concevoir de nouvelles offres de fourniture d'énergie ou de services. Ce stage s'inscrit dans le projet **Smart Customer Data**, dont l'objectif est de recueillir, valoriser et extraire des informations issues des données de consommation des clients.

Depuis une dizaine d'années, des compteurs communicants, tels que les compteurs Linky, ont été largement déployé en France et dans le monde, notamment chez les particuliers. Ces compteurs collectent la consommation totale du foyer à intervalles réguliers, produisant des séries temporelles de consommation (appelées "courbes de charge"). Ces données sont très utiles pour les fournisseurs d'électricité, par exemple pour prévoir la consommation à venir. Toutefois, comme elles agrègent la consommation de tous les appareils utilisés simultanément, il est difficile de les analyser, notamment de retrouver et d'estimer la courbe de charge individuel de chaque appareil : une information précieuse permettant d'aider les clients à réduire leur consommation. Un défi majeur consiste donc à concevoir des méthodes dites *non-intrusives*, permettant de distinguer et de quantifier la consommation de chaque appareil uniquement à partir de ces seules données agrégées.

Au cours des dernières années, la R&D d'EDF a développé diverses méthodes basées sur des approches de **Deep Learning** pour s'attaquer à cette problématique. Historiquement, ces méthodes étaient *fortement supervisées* : elles nécessitaient pour leur entraînement des données détaillées, i.e., la consommation réelle de chaque appareil à chaque pas de temps, limitant drastiquement la taille des bases de données disponibles. En effet, de telles données ne peuvent être obtenues qu'en installant des capteurs dédiés pour chaque équipement dans des foyers, une démarche coûteuse et difficile à déployer à grande échelle.

Pour faire face à cette problématique, de nouvelles approches dites *faiblement supervisées* ont récemment été investiguées. Ces approches, basé sur des méthodes de classification de série temporelle par Deep Learning, permettent de localiser les périodes précises d'utilisation de différents appareils en ne nécessitant que des **étiquettes "faibles"** (ou *weak labels*) pour être entraîné ; i.e., uniquement l'information de possession d'un appareil dans un foyer ou l'indication de son utilisation dans une large plage horaire suffit. Ces étiquettes sont beaucoup plus simples à collecter, ce qui ouvre la voie à la constitution de bases de données plus vastes, et permet in fine d'améliorer les performances des modèles grâce à la participation des utilisateurs, notamment au moyen de l'**apprentissage actif** (*Active Learning*) : une approche dans laquelle le modèle choisit les échantillons les plus incertains afin de solliciter une annotation humaine ciblée, réduisant ainsi le besoin d'étiqueter un grand volume de données.

Le but du stage est de proposer et de mettre en place une méthodes d'Active Learning afin d'améliorer les performances des modèles de désagrégations.

Le stage :

Le stage proposé consiste à investiguer et développer une méthode d'Active Learning pour améliorer les performances des algorithmes de désagrégation actuellement déployés. L'objectif est d'évaluer dans quelle mesure l'Active Learning peut :

- Améliorer la précision des méthodes existantes, en particulier pour des cas d'application spécifiques où les méthodes de désagrégation actuelles rencontrent des difficultés à les distinguer, tels que l'eau chaude sanitaire et le véhicule électrique.
- Réduire la nécessité d'étiquettes fortement supervisées.
- Permettre l'annotation à grande échelle et la personnalisation du modèle via des retours utilisateurs.

En tirant parti de labels faibles, le but est de développer des solutions plus robustes et personnalisés, tout en limitant le coût et la complexité liés à l'acquisition de données fortes.

Ce travail sera articulé en plusieurs phases.

REALISATION D'UN ETAT DE L'ART

- **Lecture et recherche d'articles scientifiques pertinents sur le sujet**
 - Etude des modèles de désagrégation : modèles développés par la R&D et existant dans la littérature.
 - Etude et recherche d'approches d'Active Learning, avec un accent sur des méthodes appliqués à des problématiques de désagrégation de courbes de charges.

Une bibliographie non-exhaustive est jointe à l'offre de stage.

- **Echange avec les chercheurs et thésards au sein de la communauté scientifique d'EDF R&D**
 - Reprise des travaux précédemment menés sur le sujet, rencontre des contributeurs principaux.
 - Familiarisation aux sujets de recherche connexes sur la courbe de charge (désagrégation de courbes de charges, classification de série temporelle pour la détection d'usages, etc.).

MISE EN PLACE ET COMPARAISON DE METHODES D'ACTIVE LEARNING

- **Implémentation d'un framework d'Active Learning utilisant les étiquettes faibles**
 - Familiarisation avec les jeux de données,
 - Prise en main des frameworks existants et développés par la R&D,
 - Proposition et développement d'une solution algorithmique.
- **Benchmark des modélisations**
 - Comparaison des performances de la solution proposée sur différents benchmarks.

RESTITUTION DES RESULTATS

- **Rédaction d'une note recensant les travaux effectués**
 - Revue de littérature commentée.
 - Conclusion sur les forces et faiblesse de la méthode proposé.
- **Rédaction du code correspondant aux travaux effectués**
 - Code bien structuré, développé en Python.
 - Code testé et documenté.

Informations complémentaires :

La R&D propose ce stage de fin d'études, à destination d'étudiants en écoles d'ingénieurs ou Master 2, spécialisés en Mathématiques Appliquées / Statistiques / Data Science / Deep Learning / IA.

Une appétence pour la recherche sera fortement appréciée.

Compétences : L'étudiant(e) sera amené(e) à mettre en œuvre et/ou acquérir des compétences en :

- Machine et Deep Learning (CNN, Transformers).
- Méthodes d'Active Learning.
- Analyse de séries temporelles (régression, classification).
- Désagrégation de la courbe de charge (a.k.a. Non-Intrusive Load Monitoring).
- Lecture et synthèses d'articles de recherche.

Le langage de programmation utilisé est Python (avec des bibliothèques telles que Pytorch, scikit-learn, etc.).

La maîtrise des bonnes pratiques Git, ainsi que la connaissance de Sphinx et Poetry/uv, sera valorisée.

Une expérience démontrée avec des modèles de Deep Learning et ces outils serait particulièrement appréciée.

Dates : Stage d'une durée de 6 mois. La date de début est flexible à partir d'avril 2025.

Lieu du stage : EDF Lab Paris-Saclay – Recherche et Développement, 7 Bd Gaspard Monge, 91120 Palaiseau.

Contacts :

Diala HAWAT (Ingénieure chercheuse, EDF), mail : diala.hawat@edf.fr

Adrien PETRALIA (Ingénieur Doctorant, EDF – Université Paris Cité), mail : adrien.petralia@edf.fr

Philippe CHARPENTIER (Ingénieur Chercheur, EDF), mail : philippe.charpentier@edf.fr

Paul BONIOL (Chercheur, INRIA – DI ENS), mail : paul.boniol@inria.fr

Guillaume HOFMANN (Chef de projet, EDF), mail : guillaume.hofmann@edf.fr

Merci d'envoyer un C.V. et une lettre de motivation sur ces e-mails.

Horaires : 35 h / semaine ; Indemnité : en fonction des formations

Bibliographie :

Les articles de recherche suivants sont pertinents pour les travaux du stage.

- A. Petralia, P. Charpentier, P. Boniol, T. Palpanas. **Appliance Detection Using Very Low-Frequency Smart Meter Time Series**. ACM International Conference on Future Energy Systems (e-energy '23), 2023.
- P. Ren, Y. Xiao, X. Chang, P. Huang, Z. Li, B. Gupta, X. Chen, X. Wang. **A Survey of Deep Active Learning**. ACM Computing Surveys (CSUR), 2021.
- T. Todic, V. Stankovic, L. Stankovic. **An active learning framework for the low-frequency Non-Intrusive Load Monitoring problem**. Applied Energy, 2023.
- G. Tanoni, E. Principi, S. Squartini. **Multilabel Appliance Classification With Weakly Labeled Data for Non-Intrusive Load Monitoring**. IEEE Transaction on Smart Grid, 2023.
- G. Tanoni, T. Sobot, E. Principi, V. Stankovic, L. Stankovic, S. Squartini. **A weakly supervised active learning framework for non-intrusive load monitoring**. Integrated Computer-Aided Engineering, 2025.
- D. Patel, A. Kumari Jain, H. Khandor, X. Choudhary, N. Batra. **Benchmarking Active Learning for NILM**. arXiv:2411.15805, 2024