

# Identification non intrusive des charges dans un bâtiment résidentiel à partir de mesures compteur à faible pas de temps

Vincent DEBUSSCHERE, Kaustav BASU, Seddik BACHA

Grenoble Alpes Université - G2ELab

**RESUME** – Nous abordons dans ces travaux l’identification non intrusive des charges des bâtiments résidentiels. L’originalité réside dans la méthode d’identification, par analyse statistique des données (algorithme de classification), qui se base sur des hypothèses réalistes restrictives sans pour autant avoir de limitation sur les modèles comportementaux des charges (variations de charges ou modèles). Ainsi, nous sommes en mesure d’identifier l’état des charges consommatrices d’énergie (et potentiellement contrôlables) en se basant uniquement sur une phase d’entraînement réduite et des mesures de puissance active agrégée sur un pas de mesure de 10 minutes, préservant donc la vie privée des habitants. Dans cette communication, après avoir décrit la méthodologie développée pour identifier les charges et leur état, ainsi que les connaissances métier fournies aux algorithmes pour augmenter l’efficacité de l’identification, nous comparons les résultats d’identification pour cinq algorithmes tirés de l’état de l’art du domaine de la classification. Ces algorithmes se différencient par leur capacité à traiter des problèmes plus ou moins complexe et se ne révèlent pas tous appropriés à tout type de charge dans le bâtiment résidentiel.

**MOTS-CLES** – Identification non intrusive, bâtiment résidentiel, classification multi-étiquette, analyse de données.

## 1 Contexte des travaux

L’identification des charges dans l’habitat (principalement résidentiel) est un domaine de recherche en plein développement, profitant des opportunités de contrôle offertes dès lors que ces méthodes d’identifications sont suffisamment précises. De nombreuses approches sont proposées dans la littérature, basées principalement sur une analyse temporelle des signaux électriques, nécessitant un taux d’échantillonnage important (la seconde ou moins) afin de détecter les variations d’état des charges. En effet, les changements d’état des charges se reflètent sur leur consommation ce qui permet de détecter leur présence dans le profil de consommation global du bâtiment. Ces méthodes sont donc basées sur la reconnaissance des signatures des charges, par analyse plus ou moins complexe des signaux électriques, de leur harmoniques ou des contenus supplémentaires que pourrait mesurer par exemple un compteur intelligent [1, 2, 3].

La principale limite de ces méthodes est la nécessité de mesures précises ou de modèles fidèles des différentes charges dans l’habitat afin de calibrer correctement les algorithmes d’identification sous peine de voir leur efficacité diminuer énormément, la période d’entraînement de ces algorithmes pouvant être même suffisamment longue pour rendre leur utilisation impraticable. Le second défaut, directement lié au premier, est l’intrusion dans l’espace privé que constitue l’habitation résidentielle. En effet, que ce soit lors du calibrage des algorithmes d’identification, puis par la quantité d’information précise recueillie sur le mode de vie des habitants, ces méthodes peuvent présenter un caractère rebutant à des utilisateurs soucieux de conserver un certain contrôle sur la connaissance que pourraient avoir des tiers de leur faits et gestes.

Pour répondre à ces questions (et tenter d’éviter ces écueils) nous proposons une approche qui ne se base pas sur de l’analyse temporelle des signaux mais sur des outils statistiques d’analyse de donnée. De plus, nous nous plaçons dans le cas d’une mesure unique d’énergie agrégée sur un pas de temps compris entre 10 et 30 minutes. Ceci nous donne plusieurs avantages. Tout d’abord, les besoins en mesure sont notablement réduits (que ce soit en volume de donnée à traiter et en complexité technique de la mesure : un compteur classique suffit). D’autre part, nous ne sommes pas en mesure d’identifier précisément le comportement des habitants, mais uniquement de reconstituer les charges de forte puissance, c’est-à-dire celles qui présentent de toute façon un intérêt réel à être contrôlées dans l’habitat. Enfin, nous avons également limité la phase d’entraînement de l’algorithme à une période très courte (deux semaines sur des données d’un an, soit moins de 4% de la base de donnée) en nous basant sur des informations très simple à obtenir sans mesures (petit tableau avec le

marquage temporel de l'utilisation ou non de certaines charges dans la maison). Ceci nous amène à une méthode quasi-non intrusive d'identification des charges dans l'habitat avec beaucoup moins de contraintes que les méthodes temporelles.

Dans cette communication, nous présentons les résultats d'identification des charges selon plusieurs algorithmes d'analyse de données, fournissant ainsi une comparaison de l'état de l'art. Ces différents algorithmes sont implémentés dans l'objectif d'intégrer dans un premier temps ces résultats dans une optimisation locale sur un triptyque habitat-production-stockage puis dans un second temps de rendre des services équivalent sur des agrégation de charges à l'échelle d'un transformateur basse tension.

Du point de vue informatique, l'originalité de ce travail est basé sur le calcul et la prise en compte d'informations spécifiques tirées de notre expérience dans le domaine du Génie Electrique fournies aux algorithmes plus classiquement appliqués dans le domaine du "text-mining" par exemple. Du point de vue plus technique, nous mettons en avant une méthodologie qui a pour avantage la prise en compte de limitations pratiques sans pour autant garder d'hypothèses de fonctionnement restrictives sur les charges pour l'aide à l'identification. Un exemple d'hypothèse couramment utilisée dans les analyses temporelles et que nous n'avons pas est par exemple le fait que deux charges ne peuvent pas changer d'état en même temps.

## 2 Principe de l'identification

Dans un premier temps, nous travaillons à identifier l'état d'une charge quelconque dans une maison à partir de l'évolution de la consommation globale, comme présenté sur l'architecture classique Figure 1(a). Cette charge est soit allumée (ON) ou éteinte (OFF). Deux remarques sont à faire ici. Tout d'abord cette simplification du problème pour une meilleure lisibilité n'exclue pas d'identifier les charge sur plus d'étiquettes que cela (qui seront alors les niveaux d'énergie consommée et non plus des états ON ou OFF). D'autre part, nous avons également appliquée cette méthode de classification pour *prédire* l'état des charges (ON ou OFF), ce qui est également un travail pouvant être généralisé à la prédiction de niveaux d'énergie [5]. L'identification et la prédiction que nous proposons se basent sur une base méthodologique similaire, mais une utilisation différente des structures algorithmiques développées (notamment le traitement et l'utilisation des données).

### 2.1 Les données

Dans le cadre du projet ANR *SuperBat*, nous disposons pour nos travaux d'une base de donnée de cent maisons dont les plus grosses charges ont été mesurées sur un pas de temps de 10 minutes pendant un an, nommée IRISE. Des données de températures et d'autres informations complètent cette base de données (nombre d'habitant par exemple) qui sera présentée plus en détail dans la communication finale.

### 2.2 Méthodologie

L'entrée principale de l'algorithme d'identification est la consommation agrégée sur un pas de temps de 10 minutes à une heure, mesurée au niveau du compteur. Lors de la phase d'entraînement, les états des charges (ON-OFF) sont fournis à l'algorithme.

La méthode utilisée dans ces travaux se base sur la classification de série de données temporelles selon une approche par fenêtres de mesures successives. L'idée est de classer les charges en fonction de leur état (ON ou OFF) ce qui permettra ensuite de les identifier en se basant uniquement sur le profil de charge global de la maison. Il s'agit donc d'une classification multi-état (ou multi-étiquettes). Le travail de classification se déroule en plusieurs étapes, qui sont décrites succinctement ci-dessous, mais qui seront explicitées dans le papier final.

1. Pré-traitement des données et groupement des charges.
2. Extraction d'information utile des données : identifier les attributs/propriétés adaptées à chaque charge.
3. Entraînement de la classification multi-étiquette.
4. Classification (donc identification) des charges.
5. Mesures évaluant la pertinence de l'identification des charges.

### 2.3 Extraction d'information utile des données

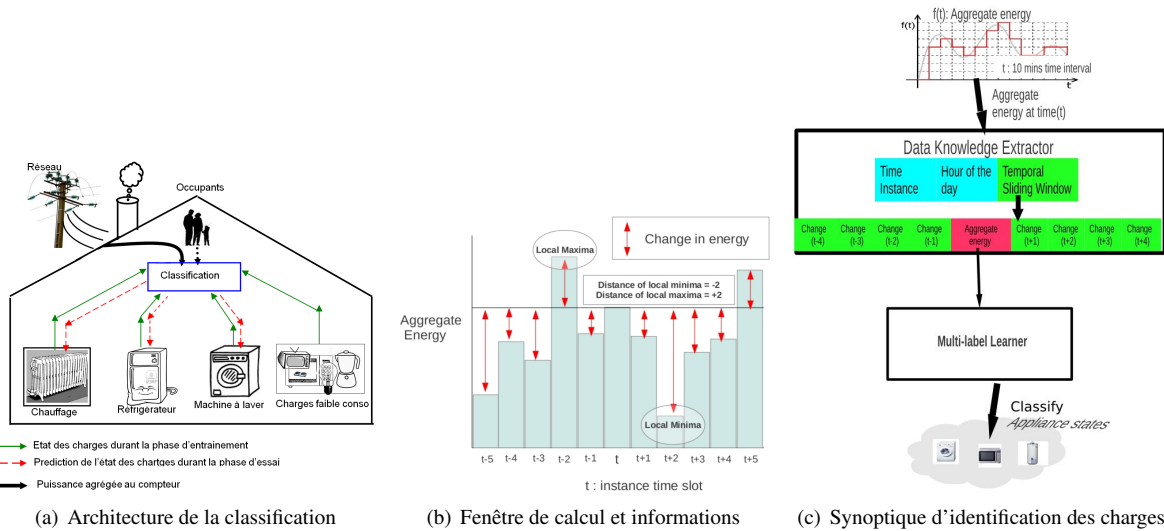
Un point clé du travail est de trouver les bonnes propriétés à faire calculer et prendre en compte à l'algorithme de classification afin qu'il soit en mesure d'identifier l'état des charges avec suffisamment de précision. Pour se faire, la courbe temporelle de la consommation globale de la maison est découpée en fenêtres successives dont le pas de temps peut varier (de 10 minutes à une heure) ainsi que la taille (en nombre d'unité ou au niveau du centrage par rapport au pas de temps analysé à chaque instance) [4]. Ce choix de pas de temps nous permet de rester dans des temps représentatifs de

remontée de mesures aux réseaux sur les prototypes de compteurs intelligents type *Linky* (les mesures en local peuvent se faire à un échantillonnage plus fin).

Pour chacune des fenêtres successives considérées, nous calculons donc plusieurs informations que l'algorithme est susceptible de lier à l'état des charges (qu'il connaît lors de la phase d'entraînement). Suivant les algorithmes, il est également en mesure de lier les états des charges entre eux. Par exemple, il y a des chances que le sèche-linge soit utilisé après la machine à laver. Cette liaison représente une information qui sortira avec plus d'importance dans les arbres de décisions de classification des différents algorithmes.

Un exemple d'information simple analysée sur une fenêtre de calcul est la variation en énergie sur les pas de temps suivants et précédents le pas de temps de l'itération considérée, comme présenté sur la Figure 1(b) où nous pouvons également positionner les maximum et minimum locaux de la fenêtre d'observation. D'autres propriétés utilisées dans nos travaux seront présentées dans le papier final.

Ainsi, chaque fenêtre d'observation successive est analysée par l'algorithme afin qu'il se construise une base de connaissance en phase d'entraînement qu'il utilisera en phase de test en ayant cette fois-ci à sa disposition uniquement la consommation globale de la maison ainsi que des informations temporelle (heure du jour, journée de la semaine). Le synoptique de l'identification est présenté Figure 1(c).



## 2.4 Algorithmes de classification

Nous avons implémenté cinq algorithmes de classification représentatifs de l'état de l'art dans le domaine. Dans le papier final, nous présenterons succinctement leurs différentes propriétés et quelles fonctionnalités les différencient entre eux. Leur nom et principale différences est résumé ci-dessous.

**LP1** : *Label powerset problem transformation* utilisant un algorithme d'arbre de décision.

**LP2** : *Label powerset problem transformation* utilisant une classification *support vector machine*.

**BR1** : *Binary relevance problem transformation* utilisant un algorithme d'arbre de décision.

**BR2** : *Binary relevance problem transformation* utilisant une classification *support vector machine*.

**MLkNN** : *Multi-label k Nearest Neighbors* avec  $K=7$ .

## 3 Résultats

Nous avons classé nos résultats en fonction de catégories de scénarios de calcul (et plus tard de topologie des maisons). Dans un premier temps, nous distinguons les cas où il y a :

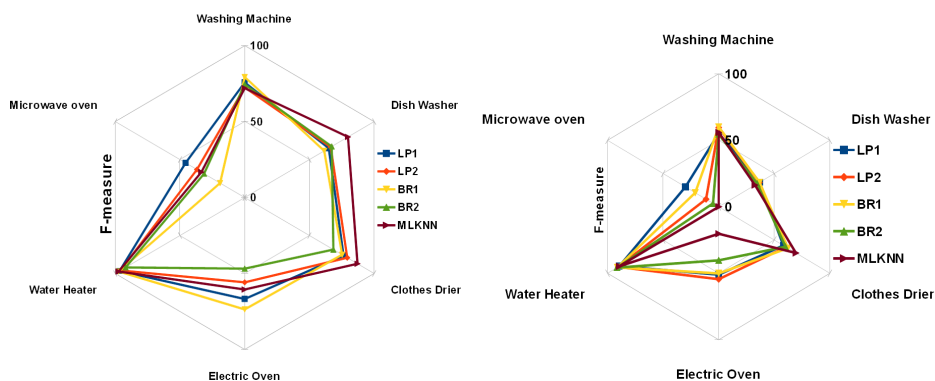
- Peu de charges de forte consommation, identifiées séparément.
- Peu de charges de forte consommation, identifiées ensemble.
- Beaucoup de charges de forte consommation avec des doubles possibles (par exemple deux télévisions).

Les résultats de classification sont exprimés en calculant un score d'identification des charges appelé *F-measure*. Celui-ci est défini comme une moyenne de deux autres indicateurs :

**Précision** : Pourcentage des états positifs (ON) correctement identifiés.

**Recal** : Rapport entre le nombre d'états positifs (ON) et le nombre total d'états positifs correctement prédits

Nous proposons Figures 1(d) et 1(e) la comparaison de cinq algorithmes dont la description est donnée en Section 2.4 en fonction de leur performance unifiée d'identification d'état des charges de fortes puissance dans une maison donnée. La Figure 1(d) est calculée pour un taux d'échantillonnage de 10 minutes alors que la Figure 1(e) pour un taux d'échantillonnage d'une heure.



(d) Identification d'état des charges (pas de 10min)

(e) Identification d'état des charges (pas de 1h)

Nous pouvons constater sur ces figures que l'identification est sensible au choix des algorithmes (notamment ceux qui sont multi-étiquette par défaut ou non) qui vont de fait présenter de meilleurs résultats en fonction des charges considérées. Nous constatons également que l'augmentation du temps d'échantillonnage a un impact fort sur les résultats d'identification, mais finalement moins important que ne pourrait le laisser penser le passage d'un échantillonnage de 10 minutes à une heure. D'autre part, l'identification se fait avec des résultats très corrects pour la plupart des charges ayant vocation à être contrôlées, même avec une mesure toutes les 10 minutes.

D'autres résultats seront proposés dans le papier final comparant les trois scénarios identifiés précédemment dans cette communication ainsi que des catégories de maisons typiques, nous amenant à considérer la méthode utilisée plutôt que les algorithmes sélectionnés.

## 4 Conclusion

Avec des informations restreintes et volontairement non-intrusive, nous sommes en mesure d'identifier les charges dans une habitation résidentielle sans avoir besoin de mesurer les variations de consommation de ces différentes charges (pas d'identification des transitions d'état requises). Les charges considérées sont celles qui consomment le plus d'énergie dans la maison, et à l'intérieur de cette catégorie, celles qui ont le potentiel d'être contrôlables, que ce soit localement ou à distance.

Nous avons comparé plusieurs algorithmes de l'état de l'art nous permettant d'atteindre ces résultats, qui sont actuellement en cours d'utilisation dans des optimisations réactive locales du triptyque bâtiment-PV-VE sous critères technico-économiques, ainsi que dans une perspective plus durable (prise en compte supplémentaire d'impacts environnementaux). Ces travaux menés à une première échelle limitée sont également développés dans un environnement d'agrégation de charges (par exemple au niveau d'un quartier) à une échelle plus générale cette fois-ci.

## Références

- [1] G. Hart, "Nonintrusive appliance load monitoring", *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.
- [2] M. E. Berges, E. Goldman, H. S. Matthews, and L. Soibelman, "Enhancing electricity audits in residential buildings with nonintrusive load monitoring", *Journal of Industrial Ecology*, vol. 14, no. 5, pp. 844–858, 2010.
- [3] M. B. Figueiredo, A. de Almeida, and B. Ribeiro, "An experimental study on electrical signature identification of non-intrusive load monitoring (nilm) systems", in *Proceedings of the 10th international conference on Adaptive and natural computing algorithms*, Berlin, Heidelberg : Springer-Verlag, 2011, pp. 31–40.
- [4] K. Basu, V. Debusschere, S. Bacha, "Load identification from power recordings at meter panel in residential households", *Electrical Machines (ICEM), 2012 XXth International Conference on*, Marseille, pp. 2098–2104.
- [5] K. Basu, V. Debusschere, S. Bacha, "Residential Appliance Identification and Future Usage Prediction from Smart Meter", *39th Annual Conference of the IEEE Industrial Electronics Society, IECON*, 2013.