

Identification et prédiction non intrusive de l'état des charges dans les bâtiments résidentiels à partir de mesures compteur à échantillonnage réduit

Vincent DEBUSSCHERE, Kaustav BASU, Seddik BACHA

Univ. Grenoble Alpes - G2ELab

RESUME – Nous abordons dans ces travaux l'identification non intrusive des charges des bâtiments résidentiels ainsi que la prédiction de leur état futur. L'originalité de ces travaux réside dans la méthode utilisée pour obtenir les résultats voulus, à savoir l'analyse statistique des données (algorithmes de classification). Celle-ci se base sur des hypothèses réalistes et restrictives sans pour autant avoir de limitation sur les modèles comportementaux des charges (variations de charges ou modèles) ni besoin de la connaissance des changements d'état des charges. Ainsi, nous sommes en mesure d'identifier et/ou de prédire 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 dix minutes, préservant donc la vie privée des habitants. Dans cette communication, après avoir décrit la méthodologie développée pour classer les charges et leurs états, ainsi que les connaissances métier fournies aux algorithmes, nous comparons les résultats d'identification pour cinq algorithmes tirés de l'état de l'art et les utilisons comme support d'application à la prédiction. Les algorithmes utilisés se différencient par leur capacité à traiter des problèmes plus ou moins complexe (notamment la prise en compte de relations entre les charges) et se ne révèlent pas tous appropriés à tout type de charge dans le bâtiment résidentiel.

Mots-Clés – Identification non intrusive, prédiction, bâtiment résidentiel, classification multi-étiquette, analyse de données.

1. CONTEXTE DES TRAVAUX

Le contrôle des charges dans l'habitat (principalement résidentiel) est un domaine de recherche en plein développement, profitant des opportunités offertes dès lors que les méthodes d'identifications et de prédiction sont suffisamment précises [3, 16]. De nombreuses approches sont proposées dans la littérature, basées principalement sur une analyse temporelle des signaux électriques et nécessitant un taux d'échantillonnage important (la seconde ou moins) afin de détecter les variations d'état des charges [6, 17]. 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 ou la connaissance a priori 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 (très) intelligent [7].

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 sous peine de voir leur efficacité diminuer. De plus, la période d'entraî-

nement de ces algorithmes peut s'avérer 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, puis par la quantité d'informations précises recueillies 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 vie quotidienne.

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 [8, 13, 4, 18]. De plus, nous nous plaçons dans le cas d'une mesure unique d'énergie agrégée sur un pas de temps compris entre dix minutes et une heure [14]. 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 mesurant la puissance active suffit) [11]. D'autre part, nous ne sommes pas en mesure d'identifier précisément le comportement des habitants, mais uniquement d'isoler l'état des charges de forte puissance (le seuil étant défini lors du paramétrage de l'algorithme), c'est-à-dire celles qui présentent de toute façon un intérêt réel à être contrôlées dans l'habitat [10]. 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 données). Tout ceci est fait en se basant sur des informations simples à obtenir sans mesures pour l'identification (il faut uniquement un marquage du temps lors de l'utilisation des charges dans la maison, sans mesures sur la charge elle-même), et avec une connaissance de l'état des charges au fil du temps dans le cas de la prédiction [9].

Ceci nous amène à une méthode quasi-non intrusive d'identification et/ou de prédiction de l'état des charges dans l'habitat avec beaucoup moins de contraintes que des méthodes temporelles [8, 5, 19]. En plus du respect de la vie privée, cette méthode va s'adapter à n'importe quel type de charge (c'est-à-dire quel que soit sa courbe de charge, ou son modèle électrique) [12]. Le modèle des charges n'ayant pas à être connu, nous n'avons pas besoin d'une mesure au niveau de la charge pour le construire. A chaque ajout ou changement de charge, ou alors sur une base régulière pour prendre en compte le vieillissement des charges et la dégradation de leur façon de consommer de l'énergie, il suffit de relancer une courte période d'entraînement de l'algorithme afin de le rendre fonctionnel sur la nou-

velle configuration de la maison [15].

Les algorithmes développés pour l'identification de l'état des charges peuvent également être utilisés pour prédire leur état futur. La différence est qu'il faut cette fois-ci une mesure au niveau des charges donnant au minimum une connaissance de leur état. Cette connaissance peut être apportée par les mêmes algorithmes, par identification, en utilisant ainsi uniquement la courbe de charge globale de la maison. Dans ce cas, l'efficacité globale de la méthode est moins bonne.

Dans cette communication, nous présentons les résultats d'identification des charges selon plusieurs algorithmes d'analyse et de classification de données, fournissant ainsi une comparaison à l'état de l'art. Puis nous appliquons les mêmes algorithmes pour réaliser de la prédiction de charge afin de montrer la viabilité de la méthode.

1.1. Les données

Dans le cadre du projet ANR SUPERBAT, nous disposons pour nos travaux d'une base de donnée, nommée IRISE, tirée d'un projet européen nommé REMODECE. Celle-ci contient le relevé de consommation, pour cent maisons et sur un an, de la quasi-totalité des charges sur un pas de temps de dix minutes. Les relevés se présentent sous la forme d'agrégation de puissance sur un pas de temps de dix minutes, donc finalement comme une représentation de l'énergie consommée sur les dix dernières minutes. Des données de températures et d'autres informations complètent cette base de données (comme le nombre d'habitant par exemple).

Les algorithmes d'identification et de prédiction présentés dans cet article sont basés sur des algorithmes de *classification*. Ceux-ci procèdent par apprentissage. Nous coupons donc la base de données en deux parties, la première pour l'apprentissage et la seconde pour la validation. La répartition dans notre cas est de 4 % pour l'apprentissage et de 96 % pour la validation. Ce choix non typique (certains algorithmes nécessitent un apprentissage sur parfois plus de 50 % de la base de donnée) est une contrainte supplémentaire que nous imposons à la méthode en vue de la rendre plus applicable à une situation réelle. En chiffres concrets, pour fournir des informations sur les charges d'un bâtiment pendant un an, l'algorithme ne nécessite qu'une période d'apprentissage de deux semaines sans mesures directe sur les charges (juste une prise de note de l'heure lors du changement d'état des charges).

Le choix de pas de temps d'une dizaine de minutes ou de l'heure est un second choix restrictif imposé dans cette étude. Celui-ci nous permet de rester dans des temps représentatifs de remontée de mesures aux réseaux sur les prototypes de compteurs intelligents (les mesures en local pouvant se faire à un échantillonnage plus fin) et de ne pas porter atteinte à la vie privée des habitants.

2. CLASSIFICATION DES CHARGES

2.1. Sélection des charges

L'objectif de ce travail est d'isoler l'état d'une charge consommant suffisamment d'énergie pour présenter un intérêt à être contrôlée dans une maison, à partir de la seule connaissance de l'évolution de la consommation globale (en puissance active) de cette maison.

Le seuil de discrimination entre les charges « intéressantes » et les autres se fait sur leur niveau de consommation d'énergie moyenne sur un pas de temps de dix minutes. Ce seuil est défini

d'une part pour faciliter l'identification des charges « utiles ». En effet, il est inutile de brouiller la courbe de consommation des charges contrôlables de forte puissance avec du bruit de mesure ou des charges consommant trop peu pour être faciles à distinguer les unes des autres. Les exemples classiques de charges de ce type sont les lampes, les chargeurs de téléphones portables, etc. Ce seuil est défini d'autre part pour mettre naturellement de côté les charges qui, même si elles étaient contrôlables, n'apporteraient pas une flexibilité suffisante sur la courbe de consommation globale pour justifier la mise en place de toute une mécanique de classification afin de détecter automatiquement leur état, ou plus tard de le prédire et le contrôler plus finement.

Ce seuil peut être choisi arbitrairement par les habitants en fonction de leur moyenne globale de consommation, ou choisi de façon plus systématique par des méthodes de groupement statistique (« clustering ») basé sur un historique de consommation initial. Contrairement à la méthode que nous présentons ici, le groupement automatique des charges en deux catégories (les charges de puissance moyenne suffisante et les autres) nécessite la connaissance de la mesure de consommation de chaque charge prise séparément. Bien que nous ayons à notre disposition les mesures individuelles des charges dans la base de donnée IRISE, nous avons choisi de rester sur un critère de seuil manuel, dans la mesure où notre méthode est forte de ne pas nécessiter ces mesures. Le seuil limite en deçà duquel les charges ne sont pas considérées a été défini sur la base de la consommation moyenne d'une télévision telle que mesurée dans les maisons de la base IRISE utilisées dans ces travaux.

2.2. Mesures et classification

L'entrée principale de l'algorithme de classification est la puissance consommée par la maison complète agrégée sur un pas de temps qui peut varier de dix minutes à une heure. Cette énergie est relevée au niveau du compteur. Lors de la phase d'apprentissage, en plus de la consommation globale et du temps, les états des charges (ON-OFF) sont fournis à l'algorithme. Lors de la phase d'utilisation, l'algorithme n'a plus à sa disposition que le temps et la consommation globale de la maison. Il donne à ce moment-là une proposition de l'état des charges, accompagné d'indices de confiances.

Les charges considérées sont soit allumées/utilisées (ON) soit éteintes/non utilisées (OFF). Il n'y a pas de mode de veille.

Deux remarques importantes sont à faire ici. Tout d'abord cette simplification du problème pour une meilleure lisibilité n'exclue pas de classer les charge sur plus d'étiquettes que cela (qui seront alors dans notre cas les niveaux d'énergie consommés sur les pas de temps considérés et non plus des états ON ou OFF). Dans le cas d'une discrétisation plus fine que juste l'état de la charge, le choix de considérer de la puissance ou de l'énergie vient directement de la mesure qui est faite sur les charges réelles, dans notre cas de l'énergie (puissance agrégée sur un pas de temps de dix minutes). D'autre part, afin d'aller plus loin que la seule connaissance de l'état de la charge, il faut impérativement une mesure de la consommation de la charge lors de la phase d'apprentissage de l'algorithme (en puissance ou énergie par pas de temps). Ceci, tout à fait possible en phase de recherche, devient plus difficile une fois que ce système est à déployer en situation réelle, ou lors d'une mise à jour du nombre de charges et de leur état. Il faut en effet installer des appareils de mesures sur chaque charge pour une durée pouvant aller jusqu'à plusieurs semaines pour de l'identification et pour la durée

d'usage pour de la prédiction.

Il est à noter que cette mesure pourrait à terme être centralisée par une « énergie box », ou un compteur de type Linky, comme présenté Figure 1. A ce stade, l'identification des charges telle que nous la présentons devient obsolète dans la mesure où les informations sur l'état des charges sont en permanence relevées ou du moins potentiellement connues par le système de management énergétique du bâtiment par une simple requête informatique. Ceci étant dit, le déploiement massif « d'énergie box » en plus de compteurs communicants (qui ne seraient pas suffisants) ne semble pas être encore une réalité à l'heure actuelle et dans tous les cas, la méthode présentée ici reste utilisable sur son aspect prédictif.

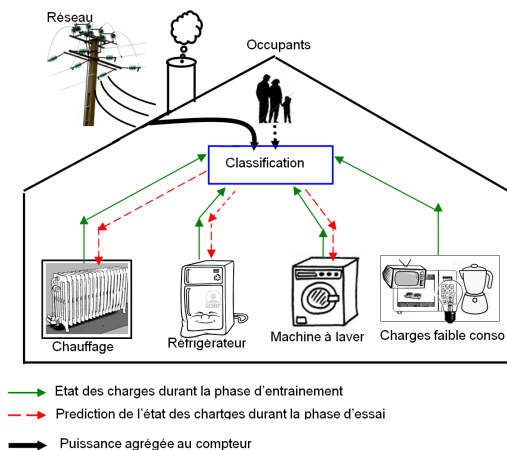


Fig. 1. Architecture de la classification par l'algorithme d'apprentissage

2.3. Principe de fonctionnement

La méthode utilisée dans ces travaux se base sur la classification de série de données temporelles selon une approche par fenêtres de mesures successives. Il s'agit 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-classe et multi-étiquettes (ou multi-état). Chaque charge de la maison prise en compte dans l'algorithme est définie comme une classe qui peut comporter deux étiquettes, ON ou OFF. Le travail de classification se déroule en plusieurs étapes, qui sont décrites succinctement ci-dessous.

1. Pré-traitement des données.
2. Calcul des attributs/propriétés pour chaque charge.
3. Apprentissage de la classification multi-étiquette.
4. Classification des charges.
5. Evaluation de la pertinence de la classification

Sans rentrer dans les détails informatiques de ces étapes, le principe de fonctionnement de l'algorithme est décrit ci-dessous.

2.4. Les étapes clés

2.4.1. Découpage de la série temporelle initiale

La classification s'opère par analyse comparative de l'évolution de séries temporelles. Il faut donc découper la série initiale S construite à partir de l'énergie consommée par le bâtiment agrégée sur des pas de dix minutes à une heure en un ensemble

de séries S_i , de longueur n , extraites de S par découpage en fenêtres successives, comme présenté, eq. (1). La taille des fenêtres peut varier (en nombre d'unité ou au niveau du centrage par rapport au pas de temps analysé à chaque instance) [1].

$$S = \{s_1, s_2, \dots, s_L\} \rightarrow \begin{cases} S_1 = \{s_1, s_2, \dots, s_n\} \\ S_2 = \{s_2, s_3, \dots, s_{n+1}\} \\ \vdots \\ S_N = \{s_{L-n-1}, s_{L-n-2}, \dots, s_L\} \end{cases} \quad (1)$$

A partir d'une série initiale S de longueur L (directement obtenue par le choix de discrétisation de la base de donnée d'un an à notre disposition) nous extrayons donc $N = (L-n)$ séries par découpage en fenêtres successives, séparées d'une unité de pas de temps. Cette unité de pas de temps a été prise de dix minutes et d'une heure pour les résultats de cet article. La longueur de la sous-série S_i , notée n est un choix qui aura une influence sur l'efficacité d'identification de l'algorithme. Il s'agit d'une valeur à déterminer au cas par cas, car il faut être capable de prendre en compte une plage assez large pour capturer des variations d'état de charge, mais pas trop pour fournir à l'algorithme le plus de séries possibles pour faciliter son travail de classification. Après plusieurs essais, nous avons choisi de travailler systématiquement dans nos travaux avec une fenêtre de découpage de dix unités, l'unité étant le pas de temps d'échantillonnage de la base de données initiale.

Lors de la phase d'apprentissage, l'algorithme parcourt la série initiale S par étapes successives (les S_i) en suivant l'information temporelle qui lui est fournie. Il prend donc en compte par construction la succession temporelle des événements. A chaque pas de temps t_i , l'algorithme choisit les sous-série correspondantes S_i . Il est possible de prendre du premier élément au dernier de cette série S_i de longueur n comme référence pour le placement du temps. Après avoir vérifié que la sensibilité de la classification des charges est faible par rapport au choix de la position de l'élément de référence dans cette sous-série, nous avons choisi de centrer la sous-série S_i sur le pas de temps t_i [1].

2.4.2. Caractéristiques supplémentaires

Pour chacune des fenêtres successives considérées, nous calculons des informations supplémentaires caractéristiques 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, s'ils sont capables de les lier.

Lors de la phase d'apprentissage, il faut donc fournir à l'algorithme, en plus de l'état des charges, du temps et de la consommation globale le plus d'informations susceptibles d'améliorer sa catégorisation des charges. Ainsi, un point clé du travail est de trouver les bonnes propriétés avec lesquels alimenter l'algorithme de classification. En effet, fournir des informations non pertinentes ne peut que dégrader l'efficacité d'identification et de prédiction. Ces informations sont calculées sur chacune des fenêtres de calculs présentées ci-dessus.

Un exemple d'information 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 2 où nous pouvons également positionner les maximum et minimum locaux de la fenêtre d'observation.

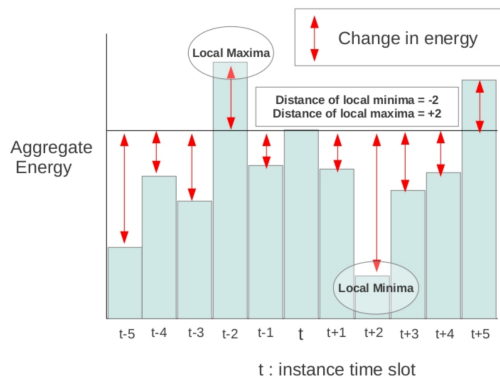


Fig. 2. Fenêtre de calcul et caractéristiques

Les informations complémentaires apportées à l'algorithme de classification représentent la valeur ajoutée de la connaissance des charges que ne peut deviner un algorithme. En plus de la position du maximum et du minimum de consommation dans la fenêtre considérée, nous ajoutons les dérivées première et seconde de cette consommation globale, des analyses statistiques sur la consommation (variation entre chaque pas de temps, moyenne et écart type). De plus, des informations temporelles supplémentaires sont données à l'algorithme : l'heure de la journée (de 0 à 23) et le jour de la semaine. Ceci permet de relever des schémas temporels sur les courbes de charge sans directement les connaître. Avec cette information, l'algorithme sera en mesure de classer les charges correspondantes lorsqu'il les reconnaîtra dans la consommation globale en phase de validation, en usage journalier ou hebdomadaire.

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 validation en ayant cette fois-ci à sa disposition uniquement la consommation globale de la maison ainsi que des informations temporelles (heure du jour, journée de la semaine). Le synoptique de fonctionnement des algorithmes de classification est présenté Figure 3.

2.5. Les algorithmes de classification

Nous avons implémenté cinq algorithmes de classification représentatifs de l'état de l'art dans le domaine. Ces algorithmes se distinguent par leur construction qui les rends naturellement ou artificiellement multi-classe. Dans le second cas, ils ne sont pas capables de lier des événements entre classes (donc entre charges pour nous). Dans ce cas, il n'est pas possible par exemple de lier l'usage du sèche-linge avec celui de la machine à laver afin d'augmenter l'efficacité de classification.

Leur nom et les caractéristiques principales des algorithmes sont résumés 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*.

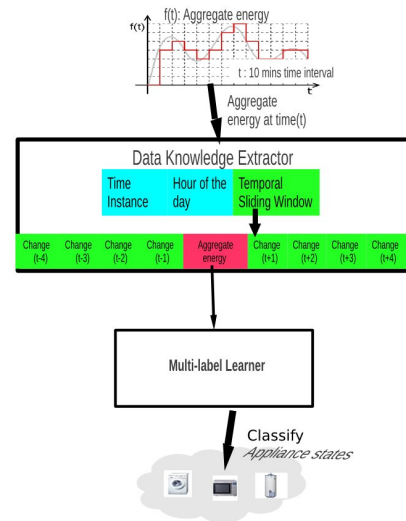


Fig. 3. Synoptique d'identification des charges

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 $K=7$.

Une description succincte des algorithmes est proposée ci-dessous [2]. Pour rappel, dans notre cas les classes sont les charges et les étiquettes (« label » en anglais) sont leurs états (ON ou OFF).

Binary Relevance, BR : Il s'agit d'une méthode de transformation de problème qui effectue un apprentissage séparé par classe sur uniquement deux étiquettes (dans notre cas ON et OFF) d'où le nom de binaire. L'algorithme BR effectue une transformation des séries initiales (par classe) en une seule série mono-classe contenant toutes les données de toutes les classes. Ensuite, l'algorithme extrait autant de tableaux qu'il y a d'étiquettes, chacun d'eux regroupant tous les attributs liés à l'étiquette.

Label Powerset, LP : L'algorithme LP considère l'ensemble des étiquettes de l'ensemble des classes comme une classe unique ayant une seule étiquette. Contrairement à l'algorithme BR, l'algorithme LP effectue son apprentissage sur un seul classificateur constitué de l'ensemble des données. L'intérêt principal est qu'il est donc possible de prendre en compte la relation entre les charges, ce que ne peut pas faire le BR. La contrepartie est le temps de calcul.

Classification : Pour ces deux algorithmes, il faut faire intervenir un second algorithme qui va classer l'apparition d'événements après transformation des données. Dans notre cas, ce sont les algorithmes « Decision Tree Learning » (DTL) et « Support Vector Machine » (SVM).

DTL est un moyen de visualiser les relations entre les informations à disposition et l'identification des charges sous la forme d'un arbre de décisions. Etant très visuel, il est systématiquement testé pour valider ses résultats.

SVM permet de grouper les charges entre elles et donc de les identifier à partir des informations mises en forme (en terme de classes et d'étiquettes) par les algorithmes LP et BR.

Multi-label k Nearest Neighbors : Cet algorithme calcule directement des distances entre séries temporelles dans l'espace qui nous convient (en fonction des valeurs d'étiquettes) permettant directement une classification et donc l'extraction d'informations sous forme d'état des charges. L'efficacité de l'identification se fera suivant le choix des paramètres constitutifs de l'algorithme. Nous avons choisi ici $k = 7$.

2.6. Performances de l'identification

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.

2.7. Prédiction

L'usage des algorithmes de classification pour effectuer de la prédiction est similaire à celui qui en est fait pour l'identification. Le principe de fonctionnement est proposé Fig. 4.

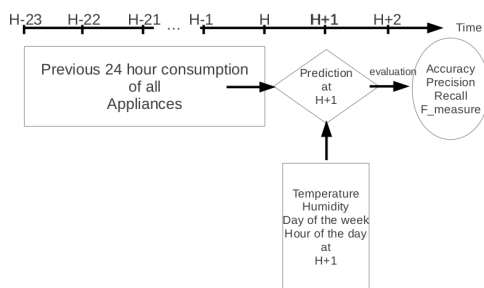


Fig. 4. Principe de fonctionnement de l'algorithme prédictif

Les algorithmes de classification sont utilisés ici pour proposer un état futur des charges en fonction de la connaissance d'un historique qui grandit à chaque instant et une phase entraînement initiale. Cela revient à identifier l'état d'une charge sans connaître encore la valeur de la consommation globale de la maison, c'est-à-dire en se basant sur des schémas similaires apparaissant dans l'historique de mesures.

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 :

1. Peu de charges de forte consommation, identifiées séparément.
2. Peu de charges de forte consommation dont certaines sont identifiées ensemble.
3. Beaucoup de charges de forte consommation avec des doubles possibles (par exemple deux télévisions).

Afin d'estimer la robustesse de l'algorithme d'identification à des configurations différentes sur le nombre de charges et leurs

caractéristiques, nous avons mené notre étude sur les trois catégories de maisons ci-dessus. Cependant, les résultats présentés ici ne concernent que la première catégorie de maison.

Les deux autres catégories de maisons sont plus difficiles à traiter pour les algorithmes de classification qui donnent des résultats moins bons. La différence avec les résultats obtenus sur la première catégorie permet de les caractériser. N'étant pas l'objet de cet article, ce ne sera pas traité ici.

3.1. Identification

Nous proposons Figures 5 et 6 la comparaison des cinq algorithmes dont la description est donnée en Section 2.5 en fonction de leur performance unifiée d'identification d'état des charges de fortes puissance dans une maison donnée. L'indicateur utilisé pour mesurer la performance est *F-measure*, présenté en Section 2.6. La Figure 5 est calculée pour un taux d'échantillonnage de dix minutes alors que la Figure 6 pour un taux d'échantillonnage d'une heure.

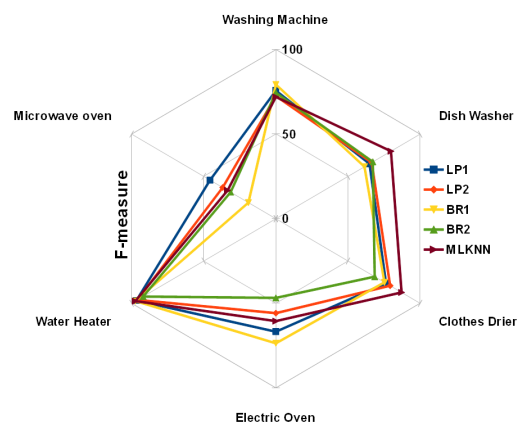


Fig. 5. Identification d'état des charges (pas de dix minutes)

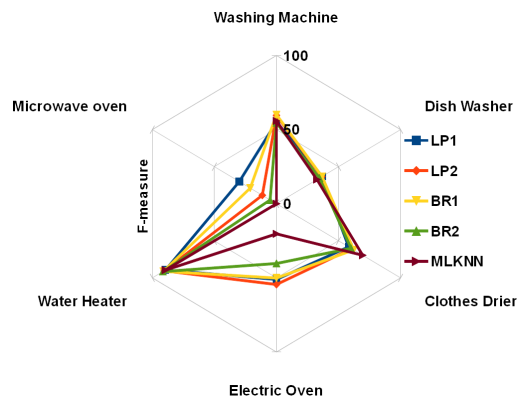


Fig. 6. Identification d'état des charges (pas d'une heure)

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 des résultats variables en fonction des charges considérées. Cependant, les différences ne sont pas aussi grandes entre algorithmes qu'entre pas de temps d'échantillonnage.

La charge la plus difficile à identifier est le four à micro-ondes. Pour les autres charges, l'algorithme MLkNN est quasiment à chaque fois le plus performant, secondé, pour les charges qui présentent un lien entre elles (sèche-linge et machine à laver) par les algorithmes multi-classe (type LP).

L'augmentation du temps d'échantillonnage a un impact significatif sur les résultats d'identification, mais finalement moins important que ne pourrait le laisser penser le passage d'un échantillonnage de dix 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 dix minutes. C'est le résultat notable de cette étude en vue d'une implémentation dans une première génération de compteurs communicants.

3.2. Prédiction

Les résultats de prédiction sont exprimés en termes de confiance, comprise entre zéro (aucune confiance dans la prédiction) et un (confiance totale dans la prédiction). Nous proposons également une représentation de la fréquence d'usage en fonction de l'heure de la journée.

Figure 7 propose les résultats de prédiction comparés aux mesures réelles, présentés sous la forme d'un diagramme fréquentiel horaire. Une prédiction est considérée comme réussie lorsque la charge considérée est dans le même état pendant les 10 minutes du pas de temps. En plus de renvoyer beaucoup d'informations aux gestionnaires réseau, ce type de graphique représente également des informations potentiellement utiles à titre informatif pour les occupants (en vue d'améliorer leur efficacité énergétique).

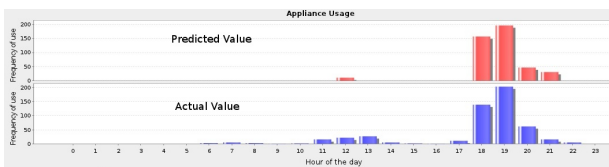


Fig. 7. Diagramme de fréquence horaire, four électrique (plaques et four)

Figure 8 propose les résultats de confiance pour la prédiction de l'usage de la télévision dans une des maisons de la base de données IRISE. Le seuil considéré pour tracer ce graphique est sévère (ON ou OFF sur 10 minutes) ce qui implique un certain nombre de ratés sur une semaine, mais reste une indication fiable sur la plage horaire d'observation.

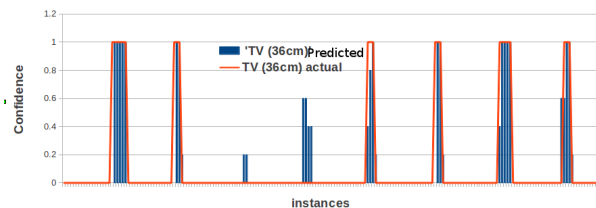


Fig. 8. Confiance de prédiction d'utilisation d'un téléviseur sur une semaine

Nous pouvons constater sur ces figures que la prédiction est viable pour les charges de forte puissance typiques que nous présentons. Ceci n'est bien sûr pas toujours le cas, notamment dans les maisons possédant plusieurs fois la même charge (deux téléviseurs par exemple). Ce point est négatif du point de vue d'un management local des charges avec comme bénéficiaire l'utilisateur (par exemple pour réduire sa facture d'électricité) mais reste pertinent du point de vue du gestionnaire réseau, car les applications visées sont plus rares, et peu dupliquées dans l'habitat (machine à laver, chauffe-eau, etc.).

3.3. Identification puis prédiction

Comme dit précédemment, l'algorithme de classification peut être utilisé pour la prédiction. La contrepartie est qu'il faut lui fournir l'état des charges les instants précédents à prédire. Deux solutions sont possibles. Soit l'état des charges est mesuré et il est utilisé directement comme entrée pour l'algorithme. Soit seule la courbe de charge est disponible, il faut donc identifier l'état des charges avec l'algorithme de classification, dont le résultat sera donné à l'étape de prédiction (basé sur les mêmes principes de classification). Cette deuxième solution a une précision plus faible, dans la mesure où les erreurs cumulées lors de l'identification impactent directement la prédiction. Son principe est proposé Fig. 9.

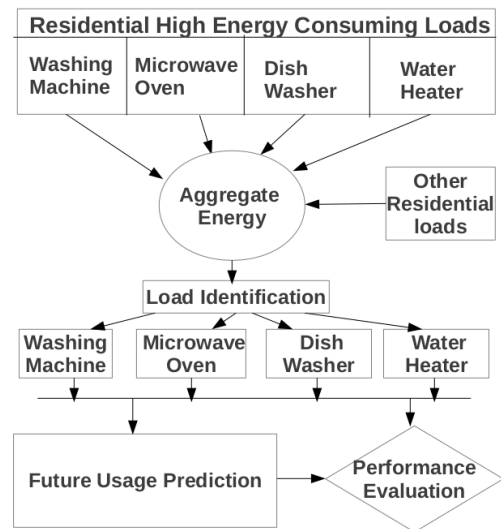


Fig. 9. Principe d'association des algorithmes d'identification et de prédiction

L'usage successif de l'identification et de la prédiction, ou la prédiction seule sont illustrés dans le tableau 1. Il y est proposé la comparaison de la prédiction de l'état des charges dans le cas d'une connaissance directe de l'état des charges ou d'une identification préalable. La différence entre les deux prédictions est directement liée aux erreurs cumulées lors de l'étape d'identification, faibles, mais inévitables.

Les résultats présentés dans le tableau 1 montrent bien la dégradation des résultats en associant deux étages de classification sans mesures réelles. Cependant, ces résultats ne sont pas non plus mauvais. De plus, ils sont variables selon les charges considérées, donc présentent tout de même un intérêt à l'usage.

Dans le cas où une mesure directe est disponible, les résultats de prédiction sont suffisants pour une utilisation réelle, ce qui valide l'intérêt de la méthode.

3.4. Limites de la méthode

La robustesse de la méthode lorsque le nombre de charges et leur caractéristiques évoluent dans le temps par rapport à la configuration initiale utilisée dans l'apprentissage n'a pas été étudiée. En principe, il est très simple de lancer à nouveau une phase d'apprentissage. D'autre part, ces méthodes ne sont pas limitées en nombre ni type de charges, car les algorithmes de classification déterminent seuls les catégories de charges et leurs étiquettes associées.

Virtuellement, la robustesse est donc très grande. D'autant plus (pour la prédiction) qu'elle est mise à jour à chaque pas

Tableau 1. Prédiction de l'état de charge avec identification préalable, ou avec mesures directes

Charges	Algorithmes	Basé sur l'identification (smart meter)			Pas d'identification (mesure directe)		
		"F-measure"	"Precision"	"Recall"	"F-measure"	"Precision"	"Recall"
Machine à laver	LP	95.11	66.66	18.66	96.58	90.74	64.42
	BR	95.13	60.51	28.22	96.61	90.00	65.57
Four à micro-onde	LP	88.22	13.33	1.41	90.47	32.83	2.75
	BR	88.27	0	0	90.40	35.92	4.62
Chauffe-eau	LP	95.71	83.42	81.68	98.73	96.29	93.29
	BR	95.96	86.16	80.33	98.73	96.29	93.29
Lave-vaisselle	LP	95.94	0	0	98.96	83.67	33.60
	BR	95.94	0	0	99.00	86.00	35.24

de temps. Cependant, elle reste à être évaluée.

4. CONCLUSION

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 utilisés 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 qui n'est pas appliquée ici est le fait que deux charges ne peuvent pas changer d'état en même temps.

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 leur variations de consommation (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 – panneaux photovoltaïques – batteries de véhicules électriques 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 en vue d'estimer un niveau de flexibilité que peut attendre un gestionnaire réseau d'un « quartier intelligent ».

D'autre part, nous avons également appliquée cette méthode de classification pour prédire l'état des charges (ON ou OFF), qui peut être lui-même généralisé à la prédiction de niveaux d'énergie [2]. L'identification et la prédiction que nous proposons ont une base méthodologique similaire, mais une utilisation différente des structures algorithmiques développées (notamment le traitement et l'utilisation des données).

L'identification et la prédiction des charges associées ensemble ont deux intérêts. Tout d'abord, elles permettent de remonter des informations décomposées par charge à l'habitant sur sa consommation, sans appareillage de mesure sophistiqué ni distribué. Il est ainsi possible d'analyser et adapter sa consommation ou détecter qu'un appareil est déficient. D'autre part, dans le cadre d'un contrôle distant, le gestionnaire du réseau de distribution pourra (suivant les évolutions futures de ses capacités de gestion) ajuster de façon plus fine la consommation à la production locale et minimiser ainsi les variations d'échanges de puissance entre différentes sections du réseau par un contrôle approprié des sources locales (qu'elles soient de production ou de consommation). Dans ce cas, l'identification des charges des bâtiments pourra se faire sur des agrégations de maisons réalisées selon un groupement préalable (statistique par exemple) basé sur des profils de consommation et d'autres informations (nombres d'habitants, région, températures, etc.)

RÉFÉRENCES

- [1] K. Basu, V. Debusschere, and S. Bacha. Load identification from power recordings at meter panel in residential households. In *Electrical Machines (ICEM), 2012 XXth International Conference on*, pages 2098–2104, Marseille, 2012.
- [2] K. Basu, V. Debusschere, and S. Bacha. Residential appliance identification and future usage prediction from smart meter. In *39th Annual Conference of the IEEE Industrial Electronics Society, IECON*, 2013.
- [3] 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*, 14 :844–858, 2010.
- [4] B.J. Birt, G.R. Newsham, I. Beausoleil-Morrison, M.M. Armstrong, N. Saldanha, and I.H. Rowlands. Disaggregating categories of electrical energy end-use from whole-house hourly data. *Energy and Building*, 50 :93–102, 2012.
- [5] M. Dong, P.C. Meira, W. Xu, and W. Freitas. An event window based load monitoring technique for smart meters. *Smart Grid, IEEE Transactions on*, 3(2) :787–796, 2012.
- [6] R.A.S. Fernandes, I.N. Da Silva, and M. Oleskovicz. Load profile identification interface for consumer online monitoring purposes in smart grids. *Industrial Informatics, IEEE Transactions on*, 9 :1507–1517, 2013.
- [7] 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*, pages 31–40, Berlin,

2011.

- [8] G.W. Hart. Nonintrusive appliance load monitoring. *Proceedings of IEEE*, 80(12) :1870–1891, 1992.
- [9] M.W. Kadous. *Temporal classification : Extending the classification paradigm to multivariate time series*. PhD thesis, The University of New South Wales, 2002.
- [10] G. Kalogridis, C. Efthymiou, S.Z. Denic, T.A., and R. Cepeda. Privacy for smart meters : Towards undetectable appliance load signatures. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, October 2010.
- [11] J.Z. Kolter, S. Batra, and Y.N.G. Andrew. Energy disaggregation via discriminative sparse coding. *Advances in Neural Information Processing Systems*, 1 :1153–1161, 2010.
- [12] W. Labeeuw and G. Deconinck. Residential electrical load model based on mixture model clustering and markov models. *Industrial Informatics, IEEE Transactions on*, 9(3) :1561–1569, 2013.
- [13] S. Laxman and P.S. Sastry. A survey of temporal data mining. *Sadhana*, 31(2) :173–198, 2006.
- [14] L.K. Norford and S.B. Leeb. Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms. *Energy and Building*, 24(1) :51–64, 1996.
- [15] T. Onoda, G. Rätsch, and K.R. Müller. Applying support vector machines and boosting to a non-intrusive monitoring system for household electric appliances with inverters, 2000.
- [16] P. Palensky and D. Dietrich. Demand side management : Demand response, intelligent energy systems, and smart loads. *Industrial Informatics, IEEE Transactions on*, 7(3) :381–388, 2011.
- [17] P. Siano. Demand response and smart grids – a survey. *Renewable and Sustainable Energy Reviews*, 30(1) :461–478, 2014.
- [18] G. Tsoumakas and I. Katakis. Multi-label classification : An overview. *International Journal of Data Warehousing and Mining*, 3(3) :1–13, 2007.
- [19] M. Zeifman and R. Kurt. Nonintrusive appliance load monitoring : Review and outlook. *Consumer Electronics, IEEE Transactions on*, 57(1) :76–84, 2011.